

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## An investigation of the mechanism of information reduction

### Thesis

#### How to cite:

Edmunds, Robert (2007). An investigation of the mechanism of information reduction. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2005 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000d4d5>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# **An investigation of the mechanism of information reduction**

Thesis submitted for the degree of Doctor of Philosophy

by

Robert Edmunds

B.Sc (Hons).

M.Sc

Faculty of Social Sciences

Department of Psychology

The Open University, UK

September 2005

AUTHOR NO U2836142  
DATE OF SUBMISSION 26 SEPTEMBER 2005  
DATE OF AWARD 26 JANUARY 2007



## **Abstract**

Contemporary theories of skill acquisition emphasise *qualitative* changes in processing as expertise is acquired (Anderson 1981; Cheng 1985; Newell 1990; 1998; Lewis 2001). In particular, describing this qualitative change as a switch from calculating the answer to the retrieval of the solutions from memory, is popular (Campitelli and Gobet, 2005; Logan 1988, 2002; Nosofsky & Palmeri, 1997; Palmeri, 1997, 1999)

Against this background, Haider and Frensch (1996, 1999a, 1999b, 2002) have recently identified some of the *quantitative* changes that may also occur with practice, changes which they term Information Reduction (IR). They demonstrated that people could ‘reduce’ to processing task-relevant segments of a stimulus, without instruction to do so. Further they found this effect was not stimulus-specific, transferring to novel item sets. This latter point was particularly troublesome for any theory reliant on the retrieval of exemplars from memory, since such a strategy will become unsuccessful for novel items.

The work presented in this thesis further explored the factors that may play a role in IR. Study 1 both replicated the basic effect and also found reduction when the visual regularity of the stimulus was varied. Study 2 realised a new target search task (TST) in which IR also developed, generalising the strategy beyond the original alphabet arithmetic task (AAT). The third study of the thesis investigated further attributes that could inhibit or facilitate reduction. The final study determined the regularity of task-redundancy necessary for IR to take place. The results are discussed in terms of the residual processing of task irrelevant items and the overall part IR must play in skill acquisition.

## **DECLARATION**

This thesis is of my own original work and is less than 100,000 words (inclusive of tables, references, and appendices). Acknowledgement has been made to all material used within this thesis where appropriate.

A doctoral research studentship from The Open University provided financial support.

The research has been presented at two conferences:

Edmunds, R., Green, A.J., and LeVoi, M.E. (2002)

Edmunds, R., Green, A.J., and LeVoi, M.E. (2003)

## **ACKNOWLEDGMENTS**

I would like to offer my gratitude to my supervisors, Dr. Alison Green, Dr. Martin LeVoi. Without their continued support and understanding this thesis would not have been possible. A sincere thanks to them for their encouragement and guidance.

I would also like to thank the staff and students of the Open University, that acted as participants in the experimental work. Once again, without their help the thesis would not have reached completion.

I am very grateful to my parents, Marjorie and Ronald for always being there for me. Particularly, I would like to thank my wife Susan for her love and understanding. She helped both bear the emotional and financial costs and listened to me talk in detail about the work. My children, Emily, Antony and Christopher, have been patient and loving even when the work made me distant from them, I appreciate their understanding.

I would like to thank Prof. John Richardson and Dr. Paula Durlach for their continued help and also Prof. Robert Snowden and Prof. Robert Johnston, who I feel started me on the road to this research.

Finally, I would like to thank my fellow colleagues and post doctoral students, Lee and Carina, who were always there to talk to during the PhD process.

## **DEDICATION**

This thesis is dedicated to my sister Gillian.

Who unknown to her, started my interest in Psychology.

# CONTENTS PAGE

Abstract.....	ii
Declaration.....	iii
Acknowledgements.....	iv
Dedication.....	v
Contents.....	vi
List of Tables.....	xi
List of Figures.....	xii
CHAPTER 1: Introduction.....	1
1.1 Thesis Structure.....	5
1.1.1 Literature Review.....	5
1.1.2 General Method section .....	6
1.1.3 Experimental Studies .....	6
1.1.4 General Discussion .....	8
CHAPTER 2: Literature Review: Skill, Attention & Learning.....	10
2.1 Introduction to the Review .....	11
2.2 Theories of Skill Acquisition .....	18
2.2.1 Fitts' Three Phase Theory .....	18
2.2.2 Schneider And Shiffrin's (1977) and Shiffrin and Schneider's (1977) Theory of Automatic and Controlled Processes.....	20
2.2.3 Anderson's Theory of Cognitive Skill Acquisition.....	24
2.2.4 Logan's (1988) Instance Theory of Automatisation .....	31
2.3 Strategy shifts and Skill Acquisition .....	45
2.3.1 Strategies in Perceptual Tasks.....	45
2.3.2 Strategies for Cognitive Skills.....	48
2.4 The Information Reduction Hypothesis, Haider & Frensch (1996, 1999a).....	50
2.4.1 A First Study of Information Reduction.....	53
2.4.2 Voluntary Strategic Processes.....	54
2.4.3 Triplet Position and Fatigue Effects.....	56
2.4.4 Processing of Irrelevant Components.....	58
2.4.5 Information Reduction and the Power Law of Practice.....	60

2.5 The Role of Attention.....	62
2.5.1 Early Selection Theories.....	62
2.5.2 Late Selection Theories.....	63
2.5.3 Space-Object Based Selection.....	65
2.6 Information Reduction and Existing Skill Acquisition Theories.....	68
2.6.1 Instance Theory, Act-R and Information Reduction.....	70
2.6.2 A Stage Model of Skill Acquisition.....	72
2.7 Implicit Learning.....	76
2.7.1 Verbal Expression of Implicit Learning.....	77
2.7.2 Implicitly Learned Environmental Regularities.....	78
2.7.3 ‘Real-World’ Implicit Learning.....	79
2.8 Future Directions.....	80
2.9 Thesis Research Questions.....	83
2.9.1. Generalising Information Reduction.....	84
2.9.2. Stimulus-Specific Components.....	84
2.9.3 Task Attributes.....	85
2.9.4 Individual Differences.....	86
2.9.5 Irrelevant Task Components.....	86
2.9.6 Conscious Awareness of the Strategy.....	87
2.10 Summary and Initial Focus.....	87
 CHAPTER 3: Overall Method.....	 90
3.1 The Alphabet Arithmetic Task (AAT).....	91
3.1.1 Measuring Information Reduction Within the AAT.....	92
3.1.2 Determining the String-Length Effect with Regression Analysis.....	93
3.1.3 Error Rate Following IR.....	94
3.2 The Target Search Task (TST).....	94
3.3 Summary of Task Variables - in the Literature.....	96
3.4 Summary of Task Variables - in the Thesis.....	99
3.5 Debriefing Questionnaires.....	101
3.6 Other Issues with the Methods used.....	102
3.6.1 Experiences of Running the Task.....	103

3.6.2 Regression Coefficients.....	105
3.6.3 Indirect Measures of IR.....	111
3.7 Practice Effects – When No IR is Possible.....	112
3.7.1 Rationale.....	112
3.7.2 Method.....	115
3.7.3 Results.....	115
3.7.4 Discussion.....	119
 CHAPTER 4: Study 1.....	120
Information Reduction in the Alphabet Arithmetic Task: Random Letter Case.....	120
4.1 Introduction.....	121
4.2 Experiment 1: Letter-Case-Alternation.....	125
4.2.1 Method.....	128
4.2.2 Results.....	131
4.2.3 Discussion.....	145
 CHAPTER 5: Study 2.....	154
Information Reduction in a Target Search Task (TST).....	154
5.1 Introduction.....	155
5.2 Experiment 2: The Target Search Task (TST).....	158
5.2.1 Target Search Tasks.....	160
5.2.2 Method.....	170
5.2.3 Results.....	174
5.2.4 Discussion.....	187
5.3 Experiment 3: Information Reduction in a Target Search task (TST): Longer Search.....	196
5.3.1 Introduction.....	196
5.3.2 Method.....	203
5.3.3 Results.....	206
5.3.4 Discussion.....	220
 CHAPTER 6: Study 3.....	234

IR and Moderating Task Attributes.....	234
6.1 Introduction.....	235
6.2 Experiment 4: Information Reduction and Task Attributes.....	241
6.2.1 Method.....	242
6.2.2 Results.....	246
6.2.3 Discussion.....	260
6.3 Individual Differences and Self-Report.....	273
6.3.1 Rationale.....	273
6.3.2 Categorising the Data.....	274
6.3.3 Individual Difference Measures.....	277
6.3.4 Results.....	285
6.3.5 Discussion.....	294
 CHAPTER 7: Study 4.....	 300
The Regularity of Source Redundancy.....	300
7.1 Introduction.....	301
7.1.1 Rationale.....	310
7.2 Experiment 5a: Pilot Study: Feedback at the End of a Block.....	313
7.2.1 Pilot Results.....	314
7.3 Experiment 5b: Main Study: Manipulating the Regularity of Source Redundancy.....	318
7.3.1 Method.....	318
7.3.2 Results.....	320
7.3.3 Discussion.....	331
 CHAPTER 8: General Discussion.....	 342
8.1 Rationale for the studies.....	346
8.1.1 Exploring Information Reduction Using the AAT and Random Letter-Case.....	346
8.1.2 Searching for Information Reduction Using a New Task.....	347
8.1.3 Examining the Extent to Which Task Attributes Might Influence Information Reduction.....	347



8.1.4 Investigating the influence of Rule Regularity on Information Reduction.....	348
8.2 Generalising the Task.....	349
8.2.1 Task Attributes.....	352
8.2.2 IR Heuristic?.....	355
8.3 Fate of Unattended Items and Residual Processing.....	356
8.3.1 Are Strategy Shifts Caused by Data-Driven or Voluntary Process...	357
8.4 Explaining Information Reduction.....	360
8.4.1 Residual Processing Hypothesis.....	360
8.4.2 IR as Vigilance.....	363
8.5 Integrating Qualitative and Quantitative of Changes that Occur During Skill Acquisition.....	364
8.5.1 Towards Integration.....	364
8.5.2 A Stage Framework.....	368
8.6 The Mechanism of IR: Implicit or Just Incidental.....	368
8.7 Practical Implications: Real-World Tasks.....	371
8.7.1 Error Avoidance.....	372
8.7.2 Conditions for IR.....	373
8.8 Implications for Future Research.....	374
8.9 Conclusions.....	376
REFERENCES.....	378
APPENDICES.....	394

## LIST OF TABLES

		Page
6.1	Analysis of First vs. Last Training Block within each condition, for Experiment 4.	250
6.2	Analysis of Last Training vs. Test Block within each condition, Experiment 4.	251
6.3	Analysis of target at trained position vs. novel position within each condition, Experiment 4.	253
6.4	Analysis of target at trained position vs. novel position within each condition, For Experiment 4: (trained position data taken from last training block).	255
6.5	Comparison of error rates between Standard and one other condition separately, Experiment 4.	256
6.6	Regression analysis results with anxiety scores as predictor and mean coefficients and error as dependant variables, for Experiment 4.	287
6.7	Logistic regression results, anxiety scores as predictors of reducer and non-reducer categories, for Experiment 4.	288
6.8	Regression analysis results with WASI scores as predictor and mean coefficients and errors as dependant variables, for Experiment 4.	289
6.9	Logistic regression results, WASI scores as predictors of reducer and non-reducer categories, for Experiment 4.	289
6.10a	Chi Square results for those that reported knowing the rule or not * 'reducer', 'non-reducer' categories, for Experiment 4	291
6.10b	Chi Square results for those that reported knowing the rule, thought they knew the rule or did not know it	291
6.11a	Chi Square results for those that reported knowing the rule or not * 'reducer', 'non-reducer' <b>error</b> categories, for Experiment 4	292
6.11b	Chi Square results for those that reported knowing the rule, thought they knew the rule or did not know it, <b>error</b> categories	293
7.1	Analysis of First vs. Last Training Block within each condition, Experiment 5b.	324
7.2	Analysis of Last Training vs. Test Block within each condition, Experiment 5b.	325

## LIST OF FIGURES

		Page
3.1	Mean regression coefficients (b & beta) from Study 1. AAT, 7-block condition.	108
3.2	Mean regression coefficients over all blocks for Target Present and Target Absent strings, for Practice Experiment.	116
3.3	Mean percent error rate* Target Position for the final, for Practice Experiment.	118
4.1	Mean Response Time (RT) for string verification over blocks for Experiment 1.	132
4.2	Mean regression coefficients over blocks; correct alphabetic strings, Experiment 1	134
4.3	Mean regression coefficients over blocks; incorrect alphabetic strings, Experiment 1	135
4.4	Mean percent error rate for the final test block, Experiment 1.	137
4.5	Mean regression coefficients for correct and incorrect alphabetic strings, Experiment 1	140
4.6.	Mean regression coefficients for correct and incorrect strings for both levels of, Experiment 1	141
4.7	Participant responses to final questionnaire, Experiment 1	144
5.1	Mean Response Time (RT) for string verification over blocks for Experiment 2.	175
5.2	Mean regression coefficients over Training and Test blocks for 1 training block condition, for Experiment 2	177
5.3	Mean regression coefficients over Training and Test blocks for 5-training block condition, for Experiment 2	178
5.4	Mean percent error rate for the final test block, Experiment 2.	180
5.5	Mean regression slopes for all conditions, target absent & target present trials, Experiment 2	183
5.6	Participant responses to final questionnaire, Experiment 2.	186
5.7	Random Walk Model illustrating upper and lower response criteria	197
5.8	Mean Response Time (RT) for string verification over blocks for Experiment 3.	207
5.9	Mean regression coefficients over Training and Test blocks for the 1 training block condition, Experiment 3	209
5.10	Mean regression coefficients over Training and Test blocks for 5-training block condition, Experiment 3.	211

5.11	Mean percent error rate for the final test block, Experiment 3.	212
5.12	Participant responses to final questionnaire for Experiment 3	217
5.13	Mean regression coefficients for the 5-training block conditions for Experiment 2 & 3: A comparison of Long and Short strings.	218
5.14	Information accumulation within a Random Walk Model	223
6.1	Example of the strings presented for each of the conditions for Experiment 4.	246
6.2	Mean Response Time (RT) for string verification over, Experiment 4	247
6.3	Mean regression slopes over Training and Test blocks for all conditions: Target absent trials, Experiment 4.	249
6.4	Mean percent error rate for the final test blocks, Experiment 4.	252
6.5	Mean percent error for target in the trained vs. novel position, for Experiment 4. (Error data for targets in the trained position taken from the last training block).	254
6.6	Mean regression slopes for all conditions over blocks compared to a baseline formed from the average of the target present, Experiment 4.	257
6.7	Participant responses to final questionnaire, for Experiment 4.	259
6.8	Mean State and Trait anxiety scores for the STAI questionnaires, for Experiment 4.	286
6.9	Mean overall Error rate and Errors for novel target positions in the final test block, participants that reported the rule or not, from Experiment 4.	293
7.1	Mean regression coefficients over training blocks for correct strings where alphabetic error occurred 100% within the triplet, for Pilot Experiment	315
7.2	Mean percent errors during test phase for correct strings, alphabetic errors post-triplet and within-triplet, for Pilot Experiment.	316
7.3	Responses to questionnaire data from Pilot Experiment	317
7.4	Mean Response Time (RT) for string verification over all blocks for the three conditions of Experiment 5b.	322
7.5	Mean regression slopes over Training and Test blocks for all conditions: Correct Strings, Experiment 5b.	323
7.6	Mean regression slopes over Training and Test blocks for all conditions: Incorrect Strings, Experiment 5b.	326
7.7	Mean percent error rate for the final test blocks, target in the trained or a novel position; correct string response errors added for comparison, Experiment 5b	327
7.8	Participant responses to final questionnaire, Experiment 5b	329

# **CHAPTER 1:**

## **INTRODUCTION**

Technology and computerisation are leading to ever more sophisticated products and interfaces. Consequently the skills and knowledge needed both in and out of work on a day-to-day basis are rapidly changing. In the past many skills could be learned by observing an expert practitioner. However, as skills become less manually and more cognitively-based the processes and knowledge underlying skilled performance may not be directly observable. Thus, watching explicit behaviour may no longer be a successful or efficient method of skill acquisition in some task domains. More frequently teaching schemes now also have to impart cognitive expertise in addition to motor skill. If these higher-level proficiencies are to be taught and carried out efficiently with the minimum risk of error, a greater understanding of the cognitions underlying skills and their acquisition is necessary.

Since skill acquisition and learning in general forms such a central part of human cognition, a long and continuing line of research has focused on this broad area. Thus, there is a great deal of literature concerning learning, skill acquisition and automaticity and a number of competing theories with this same focus. For example Anderson's (1998) ACT-R production system architecture and Logan's (1988) Instance Theory; both give comprehensive accounts of the attainment of skilled performance. These and other models are detailed further in the following literature review. What many have in common, though, is a description of qualitative changes in information processing as learning progresses. Apart from such overarching theories, there is also an extensive literature that suggests strategy selection is also part of learning, and can be accepted as part of skilled

performance (for example DeMaio, 1976; Haider & Frensch, 2002; Newton & Roberts, 2000,2001; Schunn et al 1997 and Schunn & Reder, 2001.).

While it is true that some strategy shifts results in a qualitative change in processing, there is also some evidence that quantitative changes occur too. For example Doane et al. (1999) found that strategies could be formed that reduced the number of redundant comparisons in a task, so improving performance. Gernsbacher and Faust (1991), in a study of reading comprehension skill, found that skilled readers were better able to suppress irrelevant meanings. In a study by Shapiro and Raymond (1989), training of efficient eye movements for a video game resulted in saccades only occurring when appropriate (and so to some extent only relevant) areas of the display were fully processed. Haider and Frensch (1996, 1999a, 1999b) have focused on this type of strategy change and have proposed a different view of skill acquisition, suggesting that *quantitative* changes occur in a specific way following practice at a task. Haider and Frensch's (1996) original experimental results indicated that people learned with increasing task familiarity to limit their processing to task-relevant information, ignoring any redundancy and so speeding processing and performance. They have termed this Information Reduction (IR), and suggest that it is precisely this quantitative change in processing that is under-represented in contemporary theories of cognitive skill acquisition. Of course IR would apply mostly to tasks that contain redundant elements; otherwise reduction in the information processed would lead to errors in task performance.

To set the scene, a brief synopsis of work in this area is worthwhile at this point. Haider and Frensch (1996, 1999a, 1999b, 2002) conducted a number of studies investigating IR; all

used a version of the alphabet arithmetic task (AAT). To briefly set out this paradigm, the experimental task was to verify whether strings of letters followed the alphabet or not. However, each string contained a number in brackets that was to be interpreted as the number of letters to skip in the alphabet. So for example D [4] I J K would be a correct string as 'I' follows 'D' in the alphabet when four letters (E, F, G, H) are skipped. D [4] J K L would be an incorrect string as 'J' is five letters on from 'D'. D [4] I J L would also be incorrect as 'K' is missing. In their standard version of this task the strings were constructed so that when an error occurred, it was located always at the letter immediately following the digit in brackets (for a complete description of the task and analysis methods used please refer to the Overall Methods section). Thus, only the letter-digit-letter 'triplet' was relevant to task completion and the 'post-triplet' set of letters was relegated as redundant for correct verification.

Using this method Haider and Frensch reported a number of interesting results. People were found to reduce processing to just the relevant information even if they were not informed of the redundancy. Also the strategy persisted across stimulus-sets, indicating that IR was not stimulus-specific. They trained participants on one letter-set for four blocks then switched to a novel set for the remaining blocks. The reduction that developed during the initial trials was not diminished by this change in stimuli. Further experiments (Haider & Frensch, 1999a) showed that IR was moderated by speed vs. accuracy instructions, suggesting that the strategy to attend mainly to task-relevant information may be under conscious control.

In a later study, monitoring eye fixation with a tracking device Haider and Frensch (1999b) showed that following practice participants no longer looked at the post-triplet segment. This suggested that redundancy was ignored at an early perceptual level, rather than at a later conceptual stage. Other researchers (Green & Wright, 2003) have found that IR also occurred if both sources were relevant, but provided duplicated information. So, the triplet and post-triplet segments of the strings were either *both* correct or *both* incorrect. Under these conditions people reduced to the first reliable information source they processed. Overall then, the IR effect has replicated a number of times and provided some key results. This briefly summarises the major findings; the studies are explained further in the literature review.

It could be argued that Haider and Frensch (1996, 1999a, 1999b) have provided evidence for an interesting strategy that seems at odds with contemporary theories. The IR hypothesis is perhaps more far reaching than just this though. As already mentioned, contemporary views of skilled performance do not allow for a reduction in the information processed; if IR can be shown to be a general part of everyday skilled performance, such theories must integrate the phenomenon for a full explanation. On a more practical front, in an information rich environment, many of today's tasks feature information 'overload'; IR may be one method of dealing with this. For example skilled x-ray interpreters work very quickly but with great accuracy. This is surprising, when observations reveal they often ignore abnormalities within the image. It would seem that past knowledge is brought to bear in *where* to look and what to look *for* in an image, whilst ignoring spurious information (Carmody, Nodine & Kundel, 1980). Such an account has some similarity with IR.



Brill et al. (2004) highlighted another area where information overload may take place, he reported that the army is using increasingly advanced technologies, that display amounts of information greater than the human cognitive capacity. Currently a great deal of effort, especially in military settings, has involved making information more manageable. One of these techniques is data fusion. This uses programs to gather information from many sources and highlight information pertinent to the moment within a manageable array, which seems an automated form of IR. Considering these general points, IR presents as a useful area for further research. A possible limitation though, is that, what is known about IR comes from one task, the AAT. There is obviously more that can be done to determine the antecedents of the strategy in light of this. The thesis therefore has as an aim, a further investigation of the mechanism of Information Reduction (IR) and the place of this reduction strategy in a theory of skilled performance.

## **1.1 STRUCTURE OF THE THESIS**

### **1.1.1 LITERATURE REVIEW**

The following chapter is a wide ranging literature review, not only of skill acquisition, but also other relevant literature to the main research questions. Haider and Frensch (1996, 1999a, 1999b) have drawn on very varied areas when discussing IR. The review follows their approach, and looks at attention, learning, pedagogy and skills. The reason for this, is that IR is still a hypothesis and so developing boundaries.

### 1.1.2 GENERAL METHOD SECTION

In this part of the thesis the typical alphabet arithmetic task (AAT) is described and discussed. All the various ways the task has been manipulated in the literature and for the thesis are summarised. A new target search task is also introduced and compared with the AAT. This section also describes the analyses employed, particularly regression and error analyses, methods that follow from those developed by Haider and Frensch (1996). Also included is a study to determine how much of IR can be ascribed just to general practice effects. In other words how much is due to information reduction and how much is just general speeding up following task familiarity. The analysis methods used are prone to confounding these factors, prompting this initial investigation. Finally any further problems with the paradigm are described.

### 1.1.3 EXPERIMENTAL STUDIES

Four studies were run; in each different intrinsic attributes of the tasks were manipulated.

#### *Study 1*

In the first study of the thesis the AAT, as realised by Haider and Frensch (1996), was used (as outlined on page 3). One problem with their stimuli, though, was that participants may have realised the strings were repeating and so may have attempted to learn the repeating triplet sections. So in Study 1, in order to make the string sequences ‘appear’ more unique, letter-case was randomised within each string for each trial e.g.

e (4) J

E (4) J k

E (4) j K l M n.

One surprising result in this study was evidence of participants generating mnemonics, suggesting that the triplet component was still learned. Since the AAT has been the only

task used, this learning may even be necessary for IR to develop. For the next study, a new task was developed to avoid this kind of learning.

### *Study 2*

For this studies a new target search task (TST) was realised; this allowed any number of unique strings to be generated, such as

E V W

S C J L F G

A W G Z D

The task was to determine if a letter from a memory set (C,O & V) was present in the string or not. If present, though, it was always second position from left in the string. This further investigated the generality of IR to another task, and ruled out the idea that attributes particular to the AAT may be necessary for IR to develop. A second experiment manipulated the maximum string length to determine whether increased difficulty had an effect. As suggested by Payne, Bettman and Johnson (1993) increased task complexity, such as a larger number of alternatives, will increase cognitive workload. One way to cope with this increase could be to ignore irrelevant items. The question posed was this: Does IR increase as this type of task complexity also increases?

### *Study 3*

While IR seems robust, it is possible that a number of intrinsic task attributes moderate the effect. This study was particularly aimed at investigating within one experiment the effect of string familiarity, cognitive load and source segmentation. In this case, source segmentation (e.g. the brackets) means some cue dividing the stimulus into relevant and irrelevant segments. From the previous studies it was thought that string familiarity and task difficulty would moderate IR. The effect of string segmentation was an empirical

question. Also as part of this study, individual differences such as trait/state anxiety and psychometrics (puzzle solving) were measured to determine whether these variables correlated with level of reduction in the experimental tasks.

#### *Study 4*

In this final study of the thesis one further intrinsic task variable was manipulated within the original AAT; the regularity of the string redundancy. In all the experiments thus far one part of the string during training was always consistently redundant. The obvious question is whether this 100% regularity is necessary for IR to develop. To investigate this further, conditions were created where this regularity was varied. The position of alphabetic errors was held at 100% within the triplet or 90% or 60% across training trials. IR was expected at 100%, but was to be determined at the other levels. The results of this study inform regarding the generality of IR in real world situations and also the fate of the unattended string items.

#### 1.1.4 GENERAL DISCUSSION

Following a summary of the experimental findings, the discussion aims to bring together the results in a meaningful way. It focuses on main areas such as the integration of IR with existing theories and the theoretical implications of the research beyond this. Problems with the paradigm and implications for future research are discussed. IR seems to be moderated by a number of both task intrinsic and extrinsic factors. Reduction strategies generalise to different tasks and so must be regarded as part of everyday skilled performance. There is some evidence that task redundancy does not have to be entirely consistent for a reduction strategy to be selected, which means the generality of the task is even wider.

Integrating findings with contemporary theories of skilled performance was difficult. To some degree, IR could possibly be seen as an attentional strategy that ‘guides’ what gets into an instance. Logan and Etherton (1994) suggest that the information encoded within an instance is dependent on what is attended at the time; if only relevant information becomes attended then IR may be one mechanism that ‘guides both attention and the content of an instance, a point that is expanded later for the general discussion section of the thesis.

The strategy is not item-specific though, so still presents a problem for full integration with contemporary theories of skilled performance. The studies throughout the thesis imply that the unattended items in the tasks are not so much completely ignored as checked at different degrees of frequency, and in different ways (a quick scan or a full check). The residual processing hypothesis to account for this is discussed towards the end of this section. Future research has some way to go before any full theory of the quantitative changes accompanying skill acquisition is formed. Perhaps the most interesting direction at this point would investigate how the regularity was noticed across trials. Some form of implicit learning seems likely *ad hoc*, but future study along this line may illuminate further the mechanisms of IR.

## **CHAPTER 2:**

### **LITERATURE REVIEW: SKILL, ATTENTION AND LEARNING**

*The purpose of this chapter was to review the literature relevant to skilled performance, but constrained to areas that may hold some relevance to information reduction (IR). From the review a sequence of research questions were developed.*

## **2.1 INTRODUCTION**

‘Practice makes perfect’ is a common enough proverb, accepted by the layperson with little question. Even though it is an overstatement and human performance often falls short of ‘perfect’, practice is an established route to improved or skilled performance (Anderson, 1981). Practicing a task to improve performance is a part of everyday life and is a lifelong process, the individual developing wide-ranging skills in many different domains.

When people talk of skilled performance they are usually referring to some form of motor skill, acquired through training or apprenticeship (Patrick, 1992). In the working world people are often classed as skilled, semi-skilled and unskilled. ‘Skilled’ often refers to a time served apprentice, ‘semi-skilled’ indicates some training has been received, while ‘unskilled’ suggests no formal training. In reality these boundaries are difficult to apply, as even someone employed in an unskilled role is likely to develop some special proficiency in its performance over time (Singleton, 1978). This indicates clearly that skill is a relative term that varies along a continuum; it is not something you do not have one moment, and then have the next. People can vary in their level of skill at a task and this level may change over time.

### **2.1.1 THE LAYPERSON’S VIEW OF SKILL**

The ordinary definition of skill centers on finely executed motor performance, which has as a key component ‘skill-economy’. That is, performance shows no sign of wasted effort; movement is smooth, fast, controlled and error free. In addition, the definition of the skilled worker is strongly linked to some sort of training or apprenticeship, which really is a formalised regime of practice. So, for people the concept of skill is to do with fast efficient

performance that comes about as the result of practice. This definition is applied to many areas of expertise: few would deny the skill of a concert pianist, gymnast or racing motorcyclist when they are observed performing their art. Few would also deny that these skills were the result of dedicated practice over a considerable period of time. That we accept the skill of an expert when we observe their performance is an important point, 'skill' is a hypothetical construct and as such cannot be directly observed, rather it is inferred from the level of performance (Patrick, 1992). As with any construct it is open to interpretation and indeed it is difficult to fix criteria under which someone may be labeled skilled.

Another common theme that can be drawn from the ordinary view of 'skill' is that skilled performance always has a purpose. Performance of a skill is usually goal driven. It makes no sense to be skilled unless performing the skill achieves something. The objective may be the proficiency itself; practicing just to become good at something for its own sake, such as becoming a better tennis player, but there will still be an end goal, even if that goal is self-improvement. The point of acquiring a skill is to perform the task more efficiently, thereby achieving the goal more effectively.

### 2.1.2 PSYCHOLOGICAL DEFINITIONS OF SKILL

The above suggest the layperson's view of skill as some finely tuned motor performance resulting from practice. In psychology the view of skill is far more wide ranging and is often broken down into motor, perceptual and cognitive components, though most skills contain an element of each. A somewhat more formal definition of skill comes from The Department of Employment's *Glossary of Training terms* (1971), this defines skill as;



An organized and coordinated pattern of mental and/or physical activity in relation to an object or other display of information, usually involving both receptor and *effector* processes. It is built up gradually in the course of repeated *training* or other experience. It is serial, each part from second to second is dependent on the last and influences the next. Skills may be described as perceptual, motor, manual, intellectual, social etc according to the context or the most important aspect of the skill pattern. (p.26)

This classification says a little more than the everyday concept, it suggests a range of situations where the term skill may be applied, for example, perceptual-motor would encompass driving skill, whereas intellectual/cognitive skill would describe performance at arithmetic.

Patrick (1992) suggests a taxonomy of three attributes that summarise skilled performance.

- (i) *Skills are learned or trained*, this reflects what has already been said, skills come about through formal training and also informal experience.
- (ii) *Skill implies some coordinated physical or cognitive activity to achieve a goal*, this attribute suggests that skilled performance is organized and smooth, the second part of the statement captures the goal driven nature of skill.
- (iii) *Skill implies flexible or adaptive performance*, in that the skilled performer has a large repertoire of behaviour, for example a motorcyclist may race successfully on

many different tracks and a musician expertly play a number of different pieces of music.

This implies that the skilled performer is able to select appropriate behaviour to match the task and this may extend to novel situations. However, the psychological literature does not entirely support this latter attribute: there is evidence that skills do not transfer to novel situations and high levels of training may lead to a loss of flexibility. This paradox will be returned to later in the review.

Where psychological investigation departs sharply from the general view is at the level of understanding. A description of skilled behaviour such as driving a car is not enough. Instead the mechanisms underpinning skilled behaviour are sought. Some relevant questions are

- (i) What changes as the individual moves from novice to expert?
- (ii) Can a theory of skill acquisition be provided?

### 2.1.3 VALIDITY AND UTILITY OF RESEARCH IN SKILL ACQUISITION

There are a number of theoretical and practical reasons why this is a highly important area for research.

From an academic and theoretical perspective, Anderson (1987) states that understanding how human competencies are acquired is paramount to understanding many fundamental issues in human cognition. Domains as diverse as language, problem solving and artificial intelligence (AI) all require a framework specifying the way in which 'how to' knowledge

is acquired. Further to this, understanding skill acquisition is important because it provides some constraints on theories of cognition.

The constraints imposed by such an understanding on the structure of cognitive theories serves to clarify and select amongst the many possible accounts of cognition. That is, an explanation of cognition increases in validity when it can be unified with a theory of learning. It would seem a sound framework of skill acquisition is required as a prerequisite to any theory of cognition. Indeed, Anderson's (1982, 1983) ACT\* is an overarching theory of cognitive architecture, which has as its base an account of skill acquisition.

From a more practical perspective understanding skill acquisition is important when prescribing training and is a focal area of applied psychology. Traditionally skills were learned by sitting next to a skilled practitioner and learning by observation or 'learning by doing'; however technological changes mean such methods are no longer as successful. The move away from motor-skills to more cognitive proficiencies requires a deeper understanding of expert performance if these different skills are to be passed on through effective training. There is not a great deal to be learned from *just* watching a maintenance engineer diagnosing a systems failure, as many of the skills employed are cognitive and therefore not observable. Further, the engineer may not be able to declare explicitly the knowledge that is being applied, so cannot share the knowledge verbally either. Both these problems mean that designing training for some tasks can be very problematical. Although theories of cognition are in their infancy, an analysis of skill and expertise is a necessary endeavor if training is to be improved and these problems solved.

The aim of learning a skill is to improve the performance of the individual, whether at work, in sport or everyday situations. The improvement means that less effort is required at a task and the task is usually completed more quickly. Another advantage that comes with skill or expertise is that errors are reduced. In certain situations it is the accomplishment of error free performance that is the most important. Certainly it is expected that a doctor makes as few errors as possible and confidence of an error free performance by an aircraft or nuclear power plant technician is required by all (even if not always achieved).

If errors are to be reduced to the minimum, it is necessary that a robust theory of expertise be sought. Broadbent (1990) expresses this argument when talking about avoiding catastrophes in industry. *"More generally, we can only guard against real-life errors by a theoretical understanding of the manner in which the human controller functions"* (In Patrick, 1990. p494). Often, accidents are reported in the media and the cause is ascribed to be the result of human error by skilled operators, such as errors by pilots, air traffic controllers, and nuclear power-plant operators. These are often headline news stories in the press and indicate how human performance may be error-prone even for skilled individuals.

The Three Mile Island nuclear plant incident in Pennsylvania illustrates the point, this 'event' was the result of a combination of both mechanical and human error, there is general consensus that the plant could have been brought back under control earlier if mistakes had not been made. The staff of the plant were reasonably well trained and had considerable experience with operations, so some of the errors were perhaps surprising and should be analysed further here. This extract from a narrative of events by S. Johnson (1989) describes one of the errors;

A week earlier, during a maintenance procedure, operators closed block valves (known as EFW-12 A and B) that blocked the flow of water from these emergency feedwater pumps. They were never re-opened, as they're required to be during plant operation, and none of the operators knew they were closed. An indicator light for one of the valves was covered by a yellow paper maintenance tag attached to a nearby switch, and operators simply didn't look at the others, never expecting them to be closed because they were always open during operation.

The result of this is that the pumps, running at full speed, could deliver no feedwater at all. As the operators began their checklist, the first item was "Verify emergency feed". Faust didn't see the valve indicator lights, and assumed the valves were open as they were required to be, and always had been before.

(emphasis added for the thesis)

This suggests a drawback to experience; because the valves were always previously in the required state they were not checked, when in fact they actually were one of the causes of the problem. So in this case, the operators experience worked against him and the indicator lights were not checked, with hazardous and potentially catastrophic results. Broadbent's (1990) point is well made; a theoretical understanding of skill acquisition and skilled performance is required to guard against such problems in complex environments. The

more crucial a skill is to safety, the better that skill and how it is acquired needs to be understood. This raises some important research questions, especially over the conditions in which these types of errors are likely.

A working definition of skilled performance has been provided, as has a rationale for investigating skill acquisition. This investigation has been underway from the late 1800s (Bryan & Harter, 1899), but perhaps the research took on its contemporary form in the nineteen forties, where changes in technology brought about by war required a better understanding of the facets of human performance (Welford, 1976). From this time onwards a number of theories of skill acquisition have been formed, and to some extent early principles have been retained and developed into more sophisticated accounts.

## **2.2 THEORIES OF SKILL ACQUISITION**

### **2.2.1 FITTS' THREE PHASE THEORY**

The Three Phase Theory (Fitts, 1962; Fitts & Posner, 1967) of skill acquisition was developed from experiments with pilot instructors, sports coaches and touch typists, from this work Fitts postulates three stages of skill development. <sup>1</sup> *the cognitive stage*, <sup>2</sup> *the associative phase* and finally <sup>3</sup> *the autonomous phase*. These phases are seen as sequential but not discrete with some overlapping of the phases likely to occur.

#### *The cognitive stage*

The cognitive or intellectualisation stage comes first: at this point the individual is consciously aware of rules which they rely on to perform a task. Take for example learning to drive a car: with first attempts to move forward one is very aware of operating the clutch

and trying to keep the engine speed steady with extra pressure on the accelerator. There is an explicit rule here – as the clutch pedal comes up the engine will start to stall and to compensate the accelerator pedal has to be pushed down further. When learning to drive, initially it is necessary to consciously apply such rules. In a similar way Fitts' and Posner's (1967) experiments with touch-typists showed that at first they were very aware of moving fingers to the appropriate key. For instance rules like 'move the index finger of the left hand to the right to type the letter g' were explicitly used.

### *The associative phase*

Following the cognitive phase is the associative phase, in which errors are isolated and eliminated as practice progresses. The length of this stage is dependent on the complexity of the skill and also the level of proficiency required. Fitts (1962) gives some idea of this stage being reached by typists when they know the position of the keys and how to use their fingers correctly, achieving a fair degree of speed and accuracy.

### *The autonomous phase*

As practice progresses further still, an autonomous phase is reached, according to Fitts this has two main attributes. Firstly, gradually increasing speed is achieved and for typists one keystroke every 60 msec is commonly possible (Fitts & Posner, 1967). Second, gradually increasing resistance to interference from other concurrent tasks is possible. Again using touch typists as an example, often they can carry-on an unrelated conversation whilst typing, with little loss of performance.

Perhaps the most striking quality of this stage though, is that the tasks are carried out with less and less conscious awareness and reliance on declarative rules. In fact, it is often no longer possible to verbalise the rules on which the skilled behaviour is based. Eysenck (2000) points out that while he has typed millions of words, he is now unable to tell anyone the positions of letters on the keyboard. To return to the driving example, with sufficient practice it is possible to move off and change gear whilst paying little attention to these actions, freeing resources to observe traffic conditions. To summarise, as the autonomous stage develops it becomes harder to verbalise elements of a skill and conscious attention becomes withdrawn from the task, freeing this resource to be utilised elsewhere.

These three stages seem intuitively correct, but perhaps the biggest problem is that this is a descriptive account. No real indication is made for how long the first stage will take nor how important it is. There is likely to be a great deal of variation between different tasks in the importance and duration of the cognitive stage, but no prediction of this is made in the theory. The same criticism may be levelled at the remaining stages, they are described as happening, but not when or *how* they will happen. No detail is given as to how the autonomous stage comes about (other than through practice), no mechanism for automaticity is specified, it is just accepted as a result of practice. Automaticity however, seems to be a popular and enduring concept, which has become elaborated further in contemporary theories of cognitive psychology.

### 2.2.2 SCHNEIDER AND SHIFFRIN'S (1977) AND SHIFFRIN AND SCHNEIDER'S (1977) THEORY OF AUTOMATIC AND CONTROLLED PROCESSES

Shiffrin and Schneider (1977) advanced this idea of automatic behaviour with some landmark studies. Although their theory falls under the same criticism as Fitts' in that it is



more descriptive than predictive, their experiments do provide some of the core evidence for the 'automatic vs. controlled' processing dichotomy. As already said, automaticity remains a popular way to explain improvements in performance following practice, so it is worthwhile reviewing their evidence before moving on to alternate accounts. In one of their studies (Shiffrin & Schneider, 1977), participants were to memorise a set of 1,2,3 or 4 items, either consonants or numbers. Following this 'memory set' they were presented with a display of 1,2,3 or 4 items, again either consonants or numbers. The task was to decide as quickly as possible whether any item from the memory target set was present in the display.

#### *Constant vs. varied stimulus mapping*

The crucial manipulation was the stimulus mapping between memory set and display items. In their constant mapping condition only consonants were used as memory set items and only numbers were used as distractors in the display set (or vice versa), so any item in the display set that was a consonant had to be a target. By contrast, in the varied mapping condition, both memory set and display distractors were a mixture of numbers and consonants. A large performance difference between the two mapping conditions was found. Increasing the number of items in memory and display sets had little effect in the constant mapping condition, but set size strongly affected reaction time in the varied mapping condition.

Shiffrin and Schneider (1977) interpreted this as the difference between automatic parallel processing in the constant mapping condition and serial controlled processing in the varied mapping condition. Automatic processing occurred because people have a great deal of experience discriminating letters from numbers, and so when target and distractors were

from different character sets the information could be processed quickly and in parallel. As the memory search was possible in parallel in this constant mapping condition, set size had no significant effect. This type of processing was not possible when targets and distractors were made up of mixed character sets, so a serial memory search and comparison was necessary, with more comparisons required as set size increased, and performance was slowed. To test the assumption that the effect was due to experience, a follow-up experiment extensively trained participants using only letters. The memory set was taken from the consonants 'B' through 'L' and the display distractors always taken from consonants 'Q' through 'Z'. After very many practice trials (2100) performance improved dramatically, and this was interpreted as reflecting automatic processing.

### *Structural changes*

Whilst this experimental evidence is very compelling it perhaps overemphasizes the change in processing (Patrick, 1992). The change from controlled to automatic processing is unlikely to be analogous to throwing a switch, rather it progresses along a continuum, with some processes only partially automated and under selective control (Norman & Shallice, 1980). It is also likely that automatic processes are only one aspect of skill acquisition, structural changes also accounting for improved performance as knowledge of the task is attained. Cheng (1985) has suggested that the improved performance in the studies above can be ascribed in the most part to such structural changes. As knowledge is gained of a task, different methods are employed, resulting in improved performance. She gives multiplication as an example - a structural change such as going from  $10+10+10+10+10$  to  $10*5$ , this change in method would bring about an improvement in speed.

Cheng (1985) extends this explanation to Shiffrin and Schneider (1977) studies attributing the performance in the constant mapping condition not to automatic processing but rather the selection of a particular method. Participants realised that any consonant in the display set would be a target, so just looked for any consonant amongst the numbers and did not search the memory set for a match. So really the task became one of categorization; for the constant mapping condition the memory set was given a 'category tag' and this was used to scan the display set.

This would seem to be correct, but in a reply by Schneider and Shiffrin (1985), they acknowledge the use of such a strategy in their original experiment (Shiffrin and Schneider, 1977, Experiment 3.), but suggest it does not account for all their experimental findings. It was found during set-reversals (reversing memory set for distractor set) that there was a decrement in performance. This change in performance was greater than could be explained by Cheng's (1985) categorisation hypothesis; her strategy would still be applicable even when the set-reversal had taken place. So, it would seem that some other process was disrupted: Shiffrin and Schneider (1985) propose it was a disruption in automaticity. This suggests that automaticity is a real phenomenon that cannot be easily explained by the acquisition of new strategies. However, Cheng's (1985) paper does indicate automaticity may be just one aspect of skill acquisition and other structural changes in knowledge should also be investigated. Intuitively though, automaticity is still a very important part of skilled behaviour, but so far the mechanism which underpins it has not been specified. Two theories that address this point are provided by Anderson (1982) and Logan (1988). The former is a good example of a process-based account of learning, while the latter is typical of exemplar- or instance-based theories of skill acquisition.

### 2.2.3 ANDERSON'S THEORY OF COGNITIVE SKILL ACQUISITION

Anderson's ACT (Adaptive Control of Thought) theory is more than just an account of skill acquisition, rather it is an overarching theory of the architecture of human cognition – the nature of human knowledge (Anderson, 1998). The ACT model is typical of process theories of learning, for instance it has many commonalities with the SOAR architecture (Newell, 1990; Lewis, 2001). Both attempt a complete theory of human cognition and both see knowledge represented by procedural rules that can be implemented as a computer program. SOAR and ACT do however deviate at the learning mechanism specified. SOAR has a complex mechanism termed 'chunking'. Basically, when a goal cannot be met an impasse is reached, the system then requires a sub-goal that retrieves knowledge to resolve this impasse. If successful, the knowledge leading up to the impasse and the knowledge that solved it are combined into a 'chunk' that allows the goal to be achieved in one process if it is encountered again. The learning mechanism for ACT is explained in the context of skill acquisition below.

ACT may be a unifying theory of cognition, but it is most easily understood and perhaps most plausible when applied to problem solving and skill acquisition. Anderson (1982) even suggests that problem solving is the basic mode of cognition and that skill acquisition is molded by this characteristic. So it is not surprising that the model explains these areas convincingly and in detail. ACT has gone through a number of revisions, but the core components remain with some refinement, so the following description focuses on the later ACT-R 2.0 (Anderson, 1993) and its revision ACT-R 4.0 (Anderson, 1998).

### *Declarative and procedural knowledge*

An important distinction is made within the ACT model between declarative and procedural knowledge, a distinction that fits with the outline of controlled and automatic processes outlined earlier. Declarative knowledge is 'knowing that', such as knowing that electrical resistance is measured in Ohms. This knowledge can be verbalized or written down, is explicit and so can be 'declared'. Procedural knowledge is more like 'knowing how', how to do a task such as drive a car, for instance. All or part of procedural knowledge may be implicit and not open to conscious awareness and therefore cannot be declared. In ACT-R declarative knowledge is represented as 'chunks' such as  $2+3=5$ , they are configurations of things we know. Procedural knowledge is represented as production rules, which can be general:-

IF the goal is to find the sum of  $n_1$  and  $n_2$  and  $n_1 + n_2 = n_3$

THEN say  $n_3$

Or specific:-

IF the goal is to find the sum of 2 and 3

THEN say 5

Two things can be noted from each of these production rule examples, firstly they both follow a propositional IF: Then format, as in fact do all the production rules in ACT-R (and SOAR), secondly they are both goal driven. To enable this, a hierarchal (last in first out) goal stack is the third component of the model and drives the production rules by matching to the IF part of the proposition, so the system has three types of memory, declarative, procedural and a goal stack. These modules interact dynamically and adaptively to account for learning and knowledge representation.

### *ACT and skill acquisition*

This basic outline of the ACT-R theory can be expanded to explain skill acquisition and performance, for it is the details of how productions are formed and triggered that describe the change from novice to expert. For ACT-R knowledge has to start out as declarative, so in the typist example used earlier, typists would begin by learning that to press the 'A' move the little finger on your left hand. This seems intuitively correct as normally when starting to learn a skill each stage is very consciously enacted as a number of small declarative units. Productions are then created from these chunks of knowledge by a process of compilation. In earlier models (ACT\*) this was made up of two sub-processes, composition and proceduralization, each of which progresses as the task is practiced. Composition collapsed sequences of productions that followed each other into one single production, proceduralization removed the need to retrieve declarative information into working memory when a production was enacted. However, this form of compilation and the later tuning stage (discrimination & generalization) have been abandoned in ACT-R as they created too many unwanted productions and there was little experimental support for this. At present, composition has been replaced by changes in the declarative knowledge about a task and changes in sub-symbolic process. The latter is described more fully later in this review.

In contrast SOAR retains its compositional mechanism of chunking, but it also produces far too many unwanted productions (Anderson, 1998). A different form is now proffered in ACT-R, called similarly 'production compilation'. Production compilation it seems from Anderson's (1998) description involves the setting up of a 'dependency goal', that is, a goal to understand some part of a task or instruction. So the goal is to set up to achieve a

declarative knowledge structure (set of chunks), when this goal is ‘popped’ (achieved and removed from the top of the stack) a production is induced. The substitution of variables into the production can serve to make it more general than the goal.

While this seems a more parsimonious explanation of compilation, it also lacks the clarity of Anderson’s past work. In fact Anderson (1998) has tried a number of different replacement schemes, such as learning by analogy, before settling on this most recent one. In past models, it was easier to understand how performance improved with practice, as sequences were collapsed into one production that did not require declarative knowledge to be brought into working memory. Now it seems that improvement must come from getting the correct declarative understanding of the task unit and then adapting performance by sub-symbolic processes.

What are these sub-symbolic processes? By contrast, productions and declarative knowledge are regarded as at the symbolic level, while the sub-symbolic level consists of conflict resolution – a set of processes where the computation is parallel. During conflict resolution the appropriate production is chosen to fire, out of the many that may potentially match the goal. A second process of production strength sets how fast that production will fire. Anderson suggests that these processes are really statistical estimations of probabilities and costs of a production achieving a goal, and it can be assumed that their values are set over practice. Conflict resolution can also serve to learn that a production is bad and in effect remove it from the running (a process SOAR unfortunately cannot achieve). Taken together these processes allow the system to adapt with practice, although the description is not as intuitive as the earlier models. However, it should be noted that the current model is

an evolving one; the most recent compilation mechanism is a blend of composition and proceduralization and should become clearer as work progresses (personal communication, John Anderson 24/05/2002).

To return to our typing example, it appears with regard to ACT-R, that if the goal is to obtain a chunk of knowledge on how to type the letter 'A' and the typist successfully gains that chunk, it will induce a production. The next time the goal is to type an 'A', if this production is the best match to the goal, it will fire and if successful it will become both more likely to be fired in future and will fire more quickly. Perhaps the best thing retained across models is that only knowledge that pops the goal will become a procedure; this is an advantage of learning from doing and is a key element in skill acquisition.

### *ACT and automaticity*

It was suggested earlier that ACT would provide some detail of a mechanism for automaticity; it does not have this concept as such, but Anderson (1992) suggests the phenomenon can be explained by the learning principles and conflict resolution processes within the model. As implied above the proceduralisation of productions is perhaps the most intuitive interpretation of automatic processes. 'Proceduralisation' constructs new procedures that incorporate the semantic knowledge so that it does not need to be retrieved into working memory. When this process is combined with 'composition', the result is a one step process that meets the goal efficiently and without conscious recall of declarative knowledge (Anderson, 1981).



Unfortunately ACT-R no longer has this two-stage proceduralisation and composition process in its original form, so the move over practice to fast automatic processes is a little less clear. However, Anderson (1992) posits that the most important construct in understanding automaticity is the strength of a production rule. Anderson (1982 pp170) states “....a production is automatic to the degree that it is strong”. Fortunately this concept of strengthening is still retained in ACT-R and can explain properties of automatic behaviour. For example, interference phenomena can be understood in terms of procedure strength. Anderson notes that performance is quantified by an equation, which has the multiple of production rule strength and its match to data (nominator) over the strength of competing rules and the overlap in data (denominator). So,

$$Performance \sim \frac{SAG}{NI}$$

Where  $S$  is the strength of the production rule,  $A$  is the level of activation,  $G$  is the degree to which the rule conditions are matched,  $N$  is the overlap in data elements to which competing rules match and  $I$  is the strength of those competing production rules.

As the strength of a production is increased in the nominator, there is less effect of the competing productions and data overlap in the denominator. That is, as the strength of a production increases, interference from competing elements becomes less, and the accomplishment of concurrent tasks becomes more effective. In a Stroop task, a reversal of this ratio occurs. Production rules to name the colour and the word may overlap in some elements of their conditions and so the ratio of the denominator is increased to the nominator, with a resultant drop in performance.

In a similar way the effect found in the Sternberg paradigm (Sternberg, 1969) can be understood as resulting from the formation of strong specific productions. In a Sternberg type experiment participants are given a set of items to memorise and then asked subsequently whether a probe comes from that set. As set size increases classically so does decision time, however over practice the effect of set size decreases. Anderson (1992) suggests this is because specific production rules are formed like;

‘IF: the goal is to categorise an element on the screen and 4 is on the screen

THEN: answer yes

and

‘IF: the goal is to categorise an element on the screen and 7 is on the screen

THEN: answer no.

Once productions have been formed for the memory set and foils, only one production need fire for a trial no matter how big the set size. This will only occur if mapping conditions are constant, if the memory set changes across trials, specific productions cannot be formed and the effect of set size remains. This is in agreement with the results of Sternberg-type tasks and seems a plausible explanation of the findings.

In summary, it seems Anderson’s ACT-R theory provides an account for the change in performance following practice. One caveat, though, is that ACT-R is a ‘simulation’ of human cognition; even if experimental findings are captured by the model, the brain may still utilise a different system. However, a particular strength of theories such as ACT-R and Newell’s (1990) SOAR, over everyday verbal descriptions of skill, is that they are realised as computer programs. As such, the theories on which they are based have to be

precisely specified, avoiding any vague statements or ambiguity (Stewart & West, 2006). An advantage of this, is that underlying mechanisms have to be clearly stated and not just inferred. For example, the changes in performance that come with the acquisition of skill have been described by a process that induces production rules from declarative chunks. These productions eventually allow a fast single step direct solution to a goal, without the need to retrieve declarative knowledge into working memory. Thus, the speed-up in performance following practice, can be detailed as a switch from declarative methods to the fast single-step retrieval of the correct production from memory.

Logan (1988) also proposes that practice leads to direct access of solutions from memory; in fact Logan's (1988) instance theory conceptualises automaticity as memory retrieval.

#### 2.2.4 LOGAN'S (1988) INSTANCE THEORY OF AUTOMATISATION

Logan's instance theory directly addresses the mechanism of automatization. It is an account of how automaticity is acquired that does not need to resort to a limited capacity view of attention, so side-steps the debate between modal and multiple resource theories. For Logan (1988), automaticity is viewed as memory retrieval – a single step direct access of past solutions from memory. Novices begin with an algorithm that allows a task to be completed. As they gain experience they learn specific solutions to the problem. When this is achieved, they can either respond with the algorithm or by retrieving the specific solution from memory, with enough practice they may learn to just rely on retrieval from memory. At this point Logan suggests their performance is automatic, and so is reflected as a transition from using algorithmic-based solutions to memory-based solutions.

There are three main assumptions within the theory,

1. Obligatory encoding - attending to a stimulus is sufficient to commit it to memory.
2. Obligatory retrieval - attending to a stimulus is sufficient to retrieve any associated knowledge.
3. Each encounter with a stimulus is encoded as a separate instance.

The third assumption is important as it infers a learning mechanism that relies on the accumulation of separate episodic traces with practice. So practice changes the knowledge base and this is represented as multiple episodic 'instances'. The choice between the algorithm and a solution retrieved from memory can be seen as a race: gradually retrieval of the solution from memory will overtake the algorithm and become the method of default. The concept of a race can be extended to the memory retrieval process. Each of the episodes formed over practice race against the others – the first 'past the post' will be the memory with which the individual responds. So retrieval time becomes the minimum from the distribution of the available sample, as the sample size increases it becomes likely that this minimum will decrease and performance will improve further.

### *Comparison between ACT and Instance Theory*

A learning mechanism reliant on multiple traces departs significantly from Anderson's (1998) theory, which posits that experience results in one specific production rule to achieve a task element. Although, in both the switch between a declarative and retrieval process, requires one-step retrieval of solutions from memory to be available.. This highlights another commonality between the theories: both view the goal as a solution to a problem, thus both consider skill acquisition to be framed as problem solving.

To compare the two theories further, it seems Instance theory can also account for the automaticity phenomena that were expressed above by Anderson (1992). To use the Sternberg (1969) example again, in the constant mapping condition of this experiment, as practice proceeds multiple traces could be formed for correct and incorrect categorisations. Eventually, the correct solution could be retrieved from memory without the need to use an algorithmic scan of the memory set – so the set size effect will become reduced. Logan's analogy that 'it is easy to find a wood that has a lot of trees', perhaps explains how multiple traces increase performance at this type of task. So both theories predict performance improves when a solution can be retrieved from memory in one step, but the solutions take different forms – one production or many memory traces.

Logan's (1988) explanation of the reduction in interference found during concurrent tasks following practice is also explained differently from Anderson's (1992) account. In ACT, interference is reduced as the strength of the correct production is increased in comparison to others that also partially match the stimuli. In the instance theory, interference is reduced because of the switch from algorithm to memory retrieval over practice. One task may interfere with the algorithm of another run concurrently, but when a switch is made from algorithm to retrieval, the conflicting task may no longer have a detrimental effect on what is now a memory-retrieval process. So Logan is suggesting that the switch to memory retrieval reduces the between task conflict.

To believe this, a disassociation between interference effects on the algorithm and memory process has to be accepted. However, he does not rule out the possibility that in some cases

these two different processes could interfere with each other, which indicates that practiced concurrent tasks may not always be accomplished without interference. This proposition is a rather circular argument though: there is no specification as to what properties of algorithms may conflict or what properties of memory-retrieval may be similarly affected. Only testing them empirically will inform us, so this really lacks predictive force.

Another fundamental difference between the instance and ACT theories is that in the former only the knowledge base changes, whereas ACT specifies both a change in the knowledge base and a change in underlying (sub-symbolic) processes. This suggests that ACT may be a better model of situations where people have to learn procedures that generalise across contexts and stimuli. These more general procedures may be captured by the structure of the sub-symbolic processes. By contrast, the instance theory rules out any change in how either algorithm or memories are processed. Only the knowledge base changes with practice, so there is no mechanism to capture general procedures. This leads to the specific prediction that automatised skills will not transfer to new stimuli. For transfer to occur within an instance account of learning, some form of abstraction within or across instances is required. Such abstractions are not included in Logan's (1988) theory, so stimulus-specific learning remains a limitation within Logan's explanation. This is an important point and will be returned to a number of times in this thesis.

On the other hand, Anderson's (1982) ACT theory also does not fit all task domains. Often it is apparent that people do not learn by collapsing adjacent steps (productions), rather they remember the correct solution in one attempt. So the instance theory may be the better model of tasks where people switch to retrieval without gradual 'tuning' (Logan, 1988).

This is part of the reason why this ‘staged’ method of knowledge compilation was dropped from the later ACT-R (Anderson, 1998). In effect ACT-R solutions can now become final productions in one step, so ACT-R has now moved a little closer to the instance theory. One advantage of the productions in ACT-R is that they are formalised to the point where they will run as a computer simulation, in contrast it is hard to decide completely what is an ‘instance’. First an instance is specified as an episodic trace that includes the goal, stimulus, interpretation of the two and the final response. Then it is suggested that an instance may also contain some sort of best method that can generalise to new stimuli (Logan, 1988 pp516). This immediately contradicts the earlier prediction that automatic skills are stimulus specific. So the concept of ‘an instance’ needs some further clarification.

### *Random walk models*

Logan’s (1988) instance theory has been extended to explain perceptual classification (Nosofsky & Palmeri, 1997). However, the memory retrieval assumptions of instance theory do not account for the effect of graded exemplar similarity within this domain. During classification, response times decrease as a function of similarity between the category and the exemplar. To deal with this an exemplar-based random walk (EBRW) model was proposed that borrows extensively from Logan’s original theory (Nosofsky & Palmeri, 1997, Palmeri, 1997). In this model unlike instance theory, a response does not result after the retrieval of one instance from memory. Rather all instances race to be retrieved, with rates proportional to their similarity with some presented stimuli or task. Further, again unlike Logan’s (1988) account, one instance is not enough for a response. Rather each retrieval provides incremental evidence in a random walk process.

Random walk models describe a process where information in a noisy environment is accumulated gradually until a response is made. Further, the criteria (upper and lower match/non-match boundaries) determining how much information is acquired before that response, are variables controlled by the individual. An additional variable is the starting point of the search and the relative drift rate towards one of the boundary criteria (Ratcliff, 1981, 1985). One important aspect of random walk models implied by this description is that the amount of information gathered before a response is initiated can vary. As a result all the information may not be gathered before a response, just a sufficient level determined by the parameters mentioned.

According to Palmeri's (1999) account, first with practice, it is likely instances with faster retrieval rates are added, and secondly rates are also faster the greater the similarity between exemplar and category. Accepting this, response criteria in a random walk process and specifically the EBRW model will consequently be reached more quickly following either practice or increasing exemplar-to-category similarity. Likewise, due to the nature of the random walk process, an exemplar will not need to match a category exactly before a response is elicited. All similar instances will be retrieved until sufficient information is gathered to reach a response criteria boundary. So the EBRW explains the effects of practice and also now extends instance theory to allow for similarity response grades.

The extended EBRW model still accounts for automaticity as a switch between calculation and retrieval (Palmeri, 1997) and the switch between these methods is still decided by the relative speed of each. However the model does provide one other important improvement over the original instance theory, and this is the reason it is included here. Under Logan's



(1988) theory a response is made after one instance is retrieved from memory. The limitation of this is that an instance will necessarily have to be a close match to the item presented before the initiation of an automated response. This means, in effect, that any practice improvements will be very specific to the items previously presented. In other words, transfer of the skill will be very narrow or non-existent. While this is often the case in automated performance, as said earlier skilled performance also implies some flexibility across situations.

The EBRW model however, does account for such transfer effects; Palmeri (1997) reports evidence from numerosity judgment experiments. The main experimental finding concerned the effect of novel patterns upon response times. It was found that response times in judging the number of items in a display pattern did not drop following practice, if novel patterns were only moderately distorted from those seen during training. A model with both a similarity retrieval element and a random walk process was found to fit these data adequately explaining generalised automaticity for similar but novel stimuli. The key point is that within a random walk process, only sufficient and not exact information needs to be retrieved before a response. So in this case retrieval of one exactly matching instance is not necessary for a response, rather just *sufficient* accumulated evidence from a number of similar instances. This extends Logan's (1988) original theory allowing for at least some level of transfer following automaticity. This further strengthens an instance account of skill acquisition.

### *What constitutes an instance*

Logan's (1988) instance theory appears to provide a depth of description on the cognitive changes that take place following practice at a task. However, the need to include further elements, such as random walk, so the model accounts for similarity gradients, indicates there may be a number of limitations in its explanatory power. One possible weakness in the explanation concerns what actually constitutes an 'instance'. Since this is central to the theory it should be examined in depth. This could be a very wide ranging discussion, so will focus primarily on how Logan appears to view an instance.

Logan (1988) is quite clear that an instance or processing episode is stored separately for each encounter with a stimulus, even if that encounter is identical, though separated in time. Arguably, less well defined is what each episode, captured as an instance, consists of. Logan suggests that a representation of the goal the individual is trying to achieve is encoded, along with the stimuli heeded in pursuit of that goal. Also encoded is the interpretation given to the stimulus with respect to the goal, and the response made to the stimulus. This then, covers the stimulus, context and response and leaves rather a lot of room to maneuver when one considers what could be retrieved as an instance.

Some constraint is provided by the idea that when the *same* stimulus is encountered again in the context of the *same* goal, some proportion of the episode to which it is linked is retrieved. One particular thing to note, then, is that for skilled performance to occur within instance theory, the stimulus must match one previously presented and the goal must be the same. Logan agrees that this will mean poor transfer of any practice improvements to novel

exemplars. Thus, in the original theory (Logan, 1988) an 'instance' in memory must be a one-to-one match with a previous stimulus for the execution of the skill.

Later models such as the EBRW (Palmeri, 1997) do allow this to be relaxed to a degree, but stimulus and instance must match to some extent for an automatic response to occur. While this still constrains the episode to match the stimulus in some way and tells us something of what an instance 'looks like', it does not really specify all that could possibly be part of that episode. Loose terms such as a representation of the goal, the response and an interpretation of the stimulus have also been indicated. So what gets into an episode is still not entirely specified.

The idea of encoding episodes is though a popular one and investigations into categorisation have similarly appealed to episodic or instance accounts. For example, one may judge a dog to be a dog, by comparing it to all the instances stored in memory. The example is then assigning to the category containing the most similar instances (Logan, 1988). Logan agrees that such instance theories of categorization bear a strong resemblance to his instance theory of automaticity (Logan, 1988, p515). This again contradicts an earlier description though, as now it seems that *similar* instances are retrieved to play a part in the categorisation task. This moves away from the idea that an 'instance' is a direct analog of the presented stimulus. To categorise a never-seen breed of dog, the instances retrieved could be imagined as very disparate with only certain features playing a defining role.

To some extent, this is moving towards a process view of categorisation, in which there is some abstraction across instances. Jacoby and Brooks (1984) have taken this view even

further, suggesting that the alternate views of process and episodic accounts of categorisation, may in fact, be exaggerated by the methods used to investigate each separately. They argue that the continuities between, what they term, 'analytic' and 'non-analytic' tasks, are more impressive than their discontinuities. This implies that episodes may to some degree encode or underpin abstractions. This may inform us on how instances may be employed, but still tells us little of their content.

Part of this problem may be the stimulus types chosen to investigate automaticity. Logan (1988, 1998; Logan & Etherton, 1994) used controlled lab-based tasks to investigate his instance theory, such as lexical decision, alphabet arithmetic and word-pair tasks. In these tasks, retrieval of an instance would bring to mind a single word/non-word, alphanumeric string or a pair of co-occurring words, respectively. Under these conditions it is easy to conceive of an 'instance' and to consider how, over practice, one-step retrieval from memory could provide the solution. This certainly would avoid running through some calculation or algorithm to perform the task and the choice between the two is easy to conceptualise.. However, an 'instance' of how to drive a car or ride a bicycle is much less intuitive and complex. This point is taken up by Klapp, Boches, Trabert and Logan (1991) who suggest that some tasks rely on procedural memory.

Such memories do not rely on a declarative representation and have been used to explain performance that cannot be *entirely* declared verbally; such as how to ride a bicycle (Anderson, 1981, 1982, 1983). Klapp et al (1991) go on to say that for both declarative and procedural memory, skilled performance is achieved when direct-access memory retrieval replaces the slower mechanisms that underlie performance (such as algorithms). Their

argument is it seems, that even sensory-motor skills can still be explained by instance theory. So, it would appear that an instance can also contain the 'procedure' of how to accomplish a task. This changes the view of 'what is an instance' somewhat, and at the same time further complicates the concept.

Logan certainly understands the need to specify his theory at the level of an instance and has run a series of experiments to investigate this (Lassaline & Logan, 1993; Logan, 1998; Logan & Etherton, 1994). His main findings are, (i) it seems that when attention was allocated to both items in a word-pair during a processing episode the co-occurrence of the words was encoded (Logan & Etherton, 1994). (ii) The spatial location of items may also become encoded under certain conditions (1998). (iii) Finally, Lassaline and Logan (1993) found that configural cues could also be learned as part of an instance. Within a learning episode or instance, then, information may be stored about how stimulus items are associated in a number of ways; how they go together and how they are configured. As Logan (1998) points out, instance representations are likely to be constellations of loosely connected associations, rather than a rigid snapshot-like replica of the stimulus.

Lassaline and Logan (1993) expanding this last point further, write that an instance in their experiments does not seem to preserve all of the information present in a stimulus. Rather it seems to be constrained by the nature of the task and by an attentional filter. Attributes important to the task are incorporated into the memory trace, while less important ones are not. Thus, to some extent not only are the associations between different parts of the stimulus preserved, but also the relative importance of each part. It can be concluded from

this that the instances representing processing episodes are far more than just a snap-shot of the stimulus.

Logan (1988) in his early paper on instance theory acknowledges that some element of process learning occurs, even in the tasks he was using to investigate instance learning. He proposes that this may be a different type of instance learning though, in which subjects are reminded of better ways to do a task by retrieving past instances. The idea arises again when looking at configural patterns and numerosity judgments (Lassaline and Logan, 1993). They invoke in this paper a type of memory-assisted processing mechanism; a mechanism by which memory can guide the visual scan through a configural display. In other words one can recall how best to do the task and scan the items within it. Though, they do make clear that this would not transfer to new patterns, so is still an instance account.

However, Ross (1984) provides some evidence that past episodes can be used to perform new tasks. The general idea is that the recall of past events can still act as 'reminders' to guide how to do a new task. A similar idea is proposed by Hintzman (1986), who suggests that at the point of retrieval, it is possible to abstract across similar instances to guide performance. Accepting these two views suggests firstly, that instances may possibly underpin process accounts of performance and secondly, that the influence of an 'instance' may go beyond providing an automated response to only strictly analogous stimuli.

In considering what constitutes an 'instance' the original idea of a memory trace that is a snap-shot of the original stimulus has been expounded. An instance may contain

information about the association of items within a task and also any particular spatial or configural information that may be useful. The episode may not just include declarative units, but also procedural information. It may include information on how best to do a task, though this may be in the form of 'reminding' the individual how to proceed as they abstract across similar retrieved episodes. Unfortunately, expanding the concept of an instance in this way may make the theory difficult to falsify overall, something that will become important as other ideas of skill acquisition are considered.

### *The power law of practice*

One alternative way to further evaluate Logan's (1988, 1994, 1996) and Anderson's (1982, 1993, 1998) approaches described so far, is to consider how they both map onto the power law of practice (Crossman, 1959, Newell & Rosenbloom, 1981). This 'law' proposes that the speed-up which accompanies practice follows a regular function, where initial gains are large, but diminish as practice progresses. Formally, the gains can be expressed as a power law, the simplest of which is  $T = BN^{-\alpha}$ , where  $T$  is the speed of performance,  $B$  is the speed on the first trial,  $N$  is the number of practice trials and  $\alpha$  is the learning rate (slope). When plotted as logarithmic coordinates ( $\log(T) = \log(B) - \alpha \log(N)$ ) the function becomes linear. This is important as we could expect learning to be positively exponential, whereas it is shown to have negative acceleration. This function has been shown to fit skill acquisition data so well, so often, that it has become treated as a law, subsequently, it has to be accounted for in theories of skill acquisition.

The Instance theory (Logan, 1988) explains the power function speed-up with two factors. First, thinking back to the race analogy for memory retrieval, as more traces are stored, it

becomes more likely that episodes with faster retrieval will be added and so performance will improve. The second factor is the change in probability that ever more efficient episodes will be added. That is, as the number of traces increase it becomes less likely that more efficient ones will be added, and so the increase in performance starts to slow down. This gives the negative acceleration computed by the power law. However, the reality is not quite so neat, as the change from algorithm to retrieval that is at the core of the theory, must result in a discontinuity in speed-up. Logan (1988) resolves this by pointing out any deviations from the law will be small and restricted to initial learning. Logan (1988, Expt. 4) found that there are other discontinuities in experimental data as practice progresses, these were caused by strategy shifts that aided task completion. Specifically participants used mnemonics to help with an alphabet arithmetic task. Such discontinuities are not accounted for by the instance theory, but then neither are they really explained by the ACT theory.

In comparison, the ACT (1982) theory does not map onto the power law very well, Anderson (1981) accepts that the model allows exponential speedup, but later revisions (ACT-R, 1998) suggest no alternative to indicate negative acceleration is modelled. This problem is explained away by suggesting that the power function is the result of the summation of many task sub-processes. Each sub-process may be at various stages of exponential speedup, some reaching their asymptote before others so that overall task performance follows the slower power law function. Anderson (1981, pp73-78), points to a number of different strategy shifts during certain tasks over practice, each of which could be viewed as one of the above sub-processes. This leads Anderson to conclude that there are a broad range of changes underlying improvement, some of which give exponential



gains, but sum to give an approximation of the power law. It is not clear how these strategy shifts can be incorporated into either the above theories as they capture automaticity, but strategies do seem to play a significant role in skill acquisition and so should be reviewed.

## **2.3 STRATEGY SHIFTS AND SKILL ACQUISITION**

The shift from controlled to automatic processing over practice seems to be intuitively correct and have both theoretical and experimental support. However there is also evidence that this dichotomy is not a full explanation of the performance improvements that follow practice. It would seem that with experience people also learn to employ different strategies to complete a task, and such methods also result in improved performance. As mentioned earlier, Cheng (1985) suggests that a switch from a serial memory search to a strategy of categorisation, can account for the results Shiffrin & Schneider (1977) use to support a concept of automaticity. It is probable that Cheng is correct for the most part, but it is also likely that a change of strategy combined with a move to automatic processing is a better explanation of the results (Schneider & Shiffrin, 1985). This raises the possibility that performance is improved by a combination of automatic and strategic processes.

### **2.3.1 STRATEGIES IN PERCEPTUAL TASKS**

#### *Teaching task strategies*

It appears that strategies can also be taught explicitly. Shapiro and Raymond (1989) investigated the hypothesis that efficient oculomotor behaviours can be trained and transferred to more complex visuomotor tasks, such as a 'Battle Fortress' video game. This game consisted of a space ship, which could be moved across the screen only by means of

forward power thrusts, so rotating the ship was necessary to turn or slow down. This made the ship very difficult to control. The aim of the game was to fire on a fortress and destroy it. However this could only be done if randomly appearing mines were destroyed and incoming missiles were avoided. So, the game was a complex perceptual task that required the acquisition of fine visuomotor skills. Before attempting the game participants were trained on simpler tasks, or drills. The aim of the drills was to promote either an inefficient strategy where saccades to detect peripheral display objects were induced, or an efficient strategy where attentional focus was shifted to the periphery to detect an object, but that object was not foveated. This meant that eye fixation was not shifted in a saccade; rather only attention was moved to the periphery.

The rationale for the strategic effects on game play was two-fold. First, when the visual system engages in a saccade, the system is in effect blind. So unnecessary saccades will impair performance compared to a strategy where only necessary saccades are performed. Second, learning to control the ship is better accomplished if it remains fixated. A strategy that promoted fixation on the ship while only moving attentional focus to detect peripheral objects would result in superior game performance. This is what was found: the drill that induced just shifts in attention promoted a skill that transferred to the actual game and improved play in comparison to the group trained on the inefficient strategy. This suggests two main conclusions: that successful ocular strategies that are under voluntary control can be learned, and also that these methods will transfer to other tasks with similar dimensions. Such strategies are likely to explain part of the performance improvement that comes with practicing many complex perceptual tasks.

### *Strategies in visual perception*

Further evidence that perceptual performance can be improved through the employment of a strategy was demonstrated by Snowden and Edmunds (1999). They used a 'global motion coherence' task, made up of a pattern of dots moving either in a specific direction (termed signal) or in a random direction (termed noise). Using this task, sensitivity to motion can then be determined as the percentage of signal to noise required for the detection of coherent motion. In this study (Snowden & Edmunds, 1999), the colour of the dots was also manipulated so that the noise was one colour, while the signal was another. This task was designed to tap low-level visual processes, processing that occurs before higher levels of cognition take place. So it was assumed that any advantage resulting from colour segmentation would reflect the underlying use of colour by the motion system.

An advantage was found; in fact performance was too good, correct response requiring only a very small percentage of the dots to be signal. Not only this, but the advantage was mostly negated in a subsequent experiment if *some* of the noise dots were manipulated to be the same colour as the signal. It seems unlikely that colour sensitive low-level motion processes can explain this pattern of results.

An alternative view expressed by Snowden and Edmunds (1999) to explain this anomaly, was that when signal and noise were completely segregated by colour, it became possible to apply a high-level attentional strategy to the task. For example, imagine the case where there was only one signal dot (say red) amongst many (say green) noise dots, the signal dot was likely to 'pop out'. An attentional window could then be focused on this single element and its direction detected. So long as only signal dots were red this would become a very

successful strategy requiring only one or a few signal dots for detection to be successful. This proposition was tested in two ways: first by only allowing participants to use peripheral vision and second by reducing stimulus display time. Peripheral vision lacks acuity, so if the strategy was to focus on a very small number of dots, discrimination would not be fine enough to allow motion detection. This was in fact the case and the advantage was negated. A similar finding occurred when the display time was reduced. The time needed to focus an attentional window was no longer available so the strategy failed.

Taken together this suggested that a high-level attentional strategy was used and this accounted for any advantage occurring when stimuli could be segregated by colour. Perhaps the most surprising discovery in this set of experiments was that even in a task designed to examine low-level processes, participants found strategies to employ that aid performance. It would seem the predisposition to find such strategies is pervasive in information processing, and individuals constantly search for such methods. It is a common theme in cognitive psychology that the brain is an economical information processor, and this may be one way the load is reduced. Unfortunately, in the above experiment, the issue of whether or not the strategy was under voluntary control was not investigated.

### 2.3.2 STRATEGIES FOR COGNITIVE SKILLS

In the domain of cognitive skill, there is also evidence of processing strategies following certain forms of practice. Doane and colleagues (1996) found that the difficulty of initial training influenced strategic skill acquisition. In this series of studies, participants were given a same-different judgement task, where the goal was to decide whether two polygons were identical or not. Two different polygon conditions were used, where the similarity

between target and distracter was varied. In one condition the distracters were very different to the target, which made the judgement easy, in the other condition both were very similar, so the judgement was difficult. After training on these sets, participants were transferred to discriminating between very similar novel polygons.

The main finding was that participants trained on the difficult initial discriminations using similar polygons had superior transfer performance. It was suggested that this advantage reflected the formation of different search strategies between groups. In a later study (Doane et al, 1999) extended this finding to explain more about the properties of these strategic skills. Doane suggests that for the difficult-discrimination group, an exhaustive featural comparison evolved over practice to become a constrained comparison of just relevant features. By contrast the easy-discrimination group was thought to engage in an early terminating feature comparison strategy. So in effect, the difficult-discrimination group began by comparing many of the polygon features and learned over time to disregard redundant comparisons, while the easy-discrimination group just searched until they found a difference then stopped. This indicates that the strategies were driven by the stimulus and the most effective search adopted. By manipulating the global similarity of the polygons, a strategy comparing just the overall shape was ruled out. Unfortunately though, because the polygons differed in so many dimensions, the nature of a 'constrained comparison' could not be further determined, neither could relevant and irrelevant features be specified.

Even if the exact nature of the strategies cannot be determined, it seems safe to conclude that a featural search did occur and this search differed between groups. A further study investigated the effect on feedback during the transfer task. Informing the easy-

discrimination group of their accuracy was found to have little impact and performance only improved toward the end of the session. So it would seem that, once learned, a strategy is difficult to abandon even if known to be unsuccessful. In summary, firstly, it would appear that strategies are stimulus driven, but may not be stimulus specific as transfer to novel situations can occur. Secondly, strategic knowledge also seems difficult to unlearn and the difficulty is probably a function of the amount of practice (Doane et al, 1999). Lastly, in this case a strategy was formed by the difficult-discrimination group that reduced the number of irrelevant comparisons required by the task, although the exact nature of those comparisons could not be stated.

The review so far has described predominant theories of skill acquisition. A research question suggested at this point is whether the theories described thus far completely capture the changes as skills are acquired, or are there other general aspects to skilled performance that are unaccounted for. Alternate reasons for increased performance following practice, such as strategy development seem to suggest otherwise. It is against this background that Haider and Frensch (1996) proposed their Information Reduction hypothesis, which was the research area chosen for this thesis.

#### **2.4 THE INFORMATION REDUCTION HYPOTHESIS, HAIDER AND FRENSCH (1996, 1999a)**

The theories of automaticity reviewed so far describe only qualitative changes in processing. Both Anderson's (1996) and Logan's (1988) theories describe only qualitative changes following practice, a switch from declarative to procedural processing in the former and a move from algorithmic to retrieval-processes in the latter. Even Cheng's (1985) alternative hypothesis still required the processing of the same information:

10+10+10+10+10 is still encoded even though the method changes from addition to multiplication. However, there is evidence that quantitative changes also occur and are part of skill acquisition. In Shapiro and Raymond's (1989) study, one advantage of training efficient eye movements was that saccades were only made when appropriate and so to some extent only relevant areas of the display were fully processed. This is similar to Doane and colleagues (1999) position, that a strategy can be formed that reduces the number of redundant comparisons.

Haider and Frensch (1996) have taken up this idea, and have chosen to emphasise the reduction in information processing. They posit a strategy that over time reduces the processing of irrelevant task components, which they term 'information reduction' (IR). In effect they argue that, over practice individuals learn which components are irrelevant to task completion. If these irrelevant components come to be ignored, the amount of information processed will reduce, hence information reduction (IR).

Haider and Frensch (1996) ran a series of three experiments to directly investigate this possibility. To enable the direct measurement of IR, they produced a variant of the alphabet arithmetic task (AAT). A description of this task is important here, as it remains mostly the same for all the experiments reported covering the IR hypothesis (Haider and French, 1996, 1999a., 1999b). In the task the goal is to verify alphabetic strings, which consist of a letter followed by a digit in brackets, followed by a varying number of additional letters. For example D [4] I or D [4] J K, the number in the brackets indicates the number of letters that should be skipped as the alphabet is recited. The first example string is correct as 'I' follows 'D' in the alphabet when four letters (E,F,G, & H) are skipped, whereas D [4] J K

is incorrect as 'J' does not follow when four letters are skipped. The incorrect strings were made by replacing the correct letter following the digit with the next in the alphabet. The length of the strings varied from 3 to seven symbols containing the initial letter-digit-letter triplet and a varying number of post-triplet letters.

An important aspect of their task was that while the string length could be varied, only processing of the initial letter-digit-letter triplet was required during practice for successful string verification. This was because errors always occurred in the same position, straight after the digit, within the triplet. If participants learned this abstraction as practice progressed, then gradually the post-triplet part of the string would be ignored and verification times would not increase systematically with string length. So at the start of practice there will be a string-length effect, however, as practice progresses, this effect will disappear if participants learn to restrict their attention to task-relevant components only (i.e. the triplet).

There are a number of problems with this paradigm that will be explored later in the thesis. One difficulty, at a general level, is that the AAT used by Haider and Frensch (1996), has little control over what individuals do or think as the experiment progresses. Some participants may look out for 'trick' trials (which of course occur in the test block) while other may not. In defense of this, during the learning of real-world skills there is little control over how an individual approaches a task; often different approaches will be applied to the same task. The objective of the IR studies was to determine if reduction would occur when it was not driven by instructions or cues. There is some difficulty in observing skill acquisition within the confines of an experimental setting and whilst further



controls over confounding variables may be advisable, these should not detract from the applied nature of the investigation.

#### 2.4.1 A FIRST STUDY OF INFORMATION REDUCTION

Haider and Frensch (1996, Expt.1) found that if participants knew in advance that the strings contained redundant information, they could learn to ignore the redundant part of the string, as shown by a gradual decline of the string length effect. Experiment 2 tested the hypothesis a little further in two ways. First, they no longer informed participants of the possibility of redundant information. Second, they included a final test block, where errors started to occur in the formally redundant part of the string. The results of this second experiment indicate that participants learned to reduce the information processed without any instruction that guided the formation of this strategy. Again this was shown by the decline of the systematic string length effect. Additionally, when alphabetic string errors were introduced into the post-triplet part of the string, not only did response errors increase, but also the string length effect reappeared. This strongly suggests that over practice participants had learned to ignore the redundant information, but when this strategy resulted in errors a move back to checking the whole string occurred.

##### *Transfer of information reduction*

In their third experiment (Haider & Frensch, 1996), a transfer study was conducted to determine whether the ability to ignore the redundant information was stimulus-specific. To test this, the set of strings were divided into two sub-groups, participants trained on one sub-group and were then transferred to the other. They found that the string length effect was only slightly affected by the letter-set change, suggesting that the mechanism for IR is

largely stimulus-independent. Participants appear to learn the structure of the task and are then able to transfer this to novel stimuli. However, while the set-size effect remained constant across both letter sets, overall RTs did increase on transfer. While this could have been caused by the confusion following a stimulus change, it may also indicate that there is a stimulus-specific component in the underlying mechanism/s leading to the IR effect. Two possibilities are that participants were slowed, as they were less practiced at computing the new triplets, or they may even have learned the original triplets to avoid the calculation. This is an interesting question and will be revisited later in the review.

#### 2.4.2 VOLUNTARY STRATEGIC PROCESSES

There does seem to be verification that the amount of information processed reduces over practice in the above task. One question that arises is whether this reduction is under the voluntary control or not? In a follow-up study Haider and Frensch (1999a) explored this question by manipulating task instruction. Using the same alphabet arithmetic task, their instructions to participants were to either optimise their speed or accuracy during the experiment. In addition to these two pure instruction conditions, there were two mixed instruction conditions. One, an accuracy-speed condition, where the instruction was to optimize accuracy for the first half of practice and then concentrate on speed for the second half (within subjects). The other was the reverse, a speed-accuracy condition, first an instruction to optimize speed, followed by an instruction to optimize accuracy.

For the pure speed or accuracy conditions they (Haider & Frensch, 1999a, expt.1) found the expected speed accuracy trade off, participants were faster in the speed condition and made more errors. This indicated that it was possible to change response criteria within the AAT

dependent on task instructions that were used. Importantly, they also found a difference in the string-length effect, participants stressed for speed showed a larger reduction in the string length effect than those stressed for accuracy.

Regarding the mixed conditions, the point of changing instructions mid-task, was to further test the assumption that IR was under voluntary control. If this was the case, the string length effect should be moderated by the change. To some extent Haider and Frensch (1999a) found this: the change in instruction did modify participants behaviour, but the results were not perfect. Specifically, if the manipulation worked completely, then for the accuracy-speed group, performance during the second half of practice should have resembled performance in the pure speed group. This was not really the case, there was a change, but it was far short of equivalence. The suggested reason for this was that the instructions might not have contained enough 'force' to cause a change in strategy. This seems possible, and indeed Doane (1999) suggests that once learned, even with negative feedback, a strategy can be difficult to abandon for a new method.

To explore this problem further Haider and French (1999a, Expt. 2), changed the procedure slightly so that during the speed-focused part of practice participants were also under time pressure. The rationale was that this would add additional force to the instructions and equate the performance in the speed and second half of the accuracy-speed condition. This was exactly what they found: however unexpectedly, the second phase of the speed-accuracy condition now equated to the pure accuracy condition. This is puzzling because if redundant information is isolated and ignored during the speeded part of practice, why would this irrelevant information suddenly be processed again. Even though the instruction

changed to focus on accuracy, why abandon an already successful strategy. One explanation that could fit the data is that the time constraint was removed during the accuracy stage, so maybe the extra time was taken advantage of when uncertainty was introduced by the instructional change. It would seem that there is still some clarification needed concerning the application of IR strategies. The mechanism by which people determine what information is redundant and when to ignore it is a complex issue that has only begun to be addressed.

#### 2.4.3 TRIPLET POSITION AND FATIGUE EFFECTS

One possibility is that IR was dependent on the position of the relevant information; in the AAT this was always at the start of a string. This leads to two possible reasons why an effect was found. First, in a long boring task, participants may gradually pay less attention to the whole string. This could either completely explain the IR effect or make it more likely, as it would soon be realised that success still follows when only the first part is scanned. Second, the position of the relevant information was always constant; this may be a necessary requirement for IR to occur.

To address these issues the AAT was again utilised (Haider & Frensch, 1999b. Expt.1), but the position of the relevant information, the letter-digit-letter triplet, was manipulated between conditions. In a fixed relevant-first condition, the relevant information was, as before, at the start of the string. In a fixed relevant-last condition, the triplet occurred at the end of a string and in a relevant-mixed condition the triplet occurred randomly at the start or end of a string. Again as in earlier studies, practice was followed by a transfer stage, where a percentage of the strings presented contained errors outside the triplet.

The three main results of this experiment were that first, varying the position of the task-relevant information between subjects had little effect, so it seems that relevant information does not have to appear first in the string to enable IR to occur. Second, while much reduced, IR also still happens when the position of the triplet was varied within-subjects. This suggested that while consistency in positioning aids IR, it can still occur if position varies, but the time taken to locate the relevant segment reduces the effect. Lastly, the distribution of errors during the transfer stage differed between relevant-first and relevant-last groups. Specifically, when trained on relevant-information-first, at transfer most participants initially ignored the redundant part of the string, but later came to process it. In contrast, when trained with the triplet at the end of the string, at transfer these participants mostly continued to ignore the redundant part of the string.

This was explained as an over-learned predisposition to read left to right. So the relevant-first group could not help but process some of the redundant part of the string after processing the triplet. Whereas the relevant-last group after having processed the triplet at the end would have to look from right to left and this would require extra effort. Whatever the reason for this difference it seems safe to conclude that the results rule out the possibility that the IR effect is just some expression of task fatigue, or that the position of relevant information has to remain consistent. Although this informs more of the conditions for IR, it still says little about the mechanism by which redundant information is separated from relevant information.

#### 2.4.4 PROCESSING OF IRRELEVANT COMPONENTS

On a somewhat different tack, another problem was that the string-length effect was only an indirect measure of IR. It cannot safely be assumed that the redundant part of the strings receives greatly attenuated processing. Haider and French (1999b) attempt to address this issue by measuring participants' eye movements as they undertake the alphabet arithmetic task (AAT). By determining on which parts of the string people fixated, a more direct measure of IR could be realised. The rationale was that if participants fixated only on task-relevant information, IR occurred at an early perceptual level. However, if both task-relevant and task-irrelevant parts of the string were fixated, IR must have taken place at a higher cognitive level. That is, the redundant information was perceived, but was suppressed at a later conceptual stage. Again using the AAT and the relevant-first and relevant-last conditions laid out above there were two main results.

First, for the relevant-last condition, participants behaved in complete accordance with the acquisition of an information reduction strategy. They began by serially attending to all of the string, but as practice progressed the string-length effect reduced and fixation changed. No longer was task-redundant data fixated, instead only task-relevant information became the focus of fixation. Those participants that did not show a reduction in the string-length effect (non-reducers) continued to fixate on all the string. This leads to the conclusion that IR occurs at a perceptual level and redundant information becomes unattended very early on in processing. Unfortunately this pattern was not completely replicated by the relevant-first condition: some fixation still occurred on the redundant part of the string. This counters the earlier conclusion and the only explanation offered is that this weaker pattern was caused again by an over-learned left-right reading habit.

While the suggestion of an over-learned habit may appear plausible, if the actual stimuli are considered, it seems more problematic. Each letter was 3.8cms and letters were 6cms apart, for the seven item strings the display was at least 68.6 cms wide. What this means is that to scan the redundant part of the string considerable eye movement was required. So not reducing would be effortful, even if the habit were to read from left to right. So perhaps it is surprising that in this modified AAT, the two groups were not equivalent. Indeed the task is so changed by increasing the display size, that the effect even for the relevant-last group cannot be assumed to generalise to the standard task that covers a much smaller visual angle and would require far less, if any, eye movement.

Posner (1980) indicates that peripheral areas of the visual field may undergo substantial processing, and that attention can be disassociated from the point of fixation to the periphery. If this view is accepted, a measure of fixation is unlikely to rule out that some part of the redundant information is processed. Further support for this opinion comes from theories of focal attention such as the 'zoom lens' analogy (Eriksen & Yei-Yu Yeh, 1985, Eriksen & St.James, 1986). Focal attention in this case takes the form of a lens where there is a trade-off between resolving power and field of view. The zoom can be narrowed with increased detailed serial processing or widened with parallel dispersed processing. One aspect of this model relevant to the present discussion is that there is no step-like cut-off in attention between the areas inside and outside of focus. Rather there is a gradual drop-off of processing from one area to another. So it seems unlikely that the redundant part of the AAT strings receive no processing, although that processing may degrade for items further into the post-triplet part of the stimulus. Overall, it appears possible that attention will span

beyond the triplet or point of fixation, even after IR is acquired, although the qualities of that processing is an empirical question.

#### 2.4.5 INFORMATION REDUCTION AND THE POWER LAW OF PRACTICE.

As written earlier in the review, one way to evaluate any theory of skill acquisition is to look at how it maps onto the power law of practice (Crossman, 1959, Newell & Rosenbloom, 1981). Haider and Frensch (2002) did just this with IR, re-analysing their previous AAT data. The main thrust of their argument was that there was some contradiction between their data and what would be expected from the power law. According to the latter, practice results in a *continuous* negatively accelerated curve. However, their findings were that IR results in a discontinuous change over practice. That is, they have evidence that when a switch to an IR strategy takes place, this causes a discontinuity, a sudden speed-up.

How can these two views be reconciled? Haider and Frensch (2002) following others (Delaney, Reder, Staszewski & Ritter, 1998; Palmeri, 1999 and Rickard, 1997) report an ingenious explanation to account for the contradiction. At its core, their argument was that while discontinuities do occur for practice at some tasks, they take place at different times for different individuals. When these temporally disparate steps in performance are aggregated, the curve is 'smoothed out' to give the typical continuous power-function.

It is possible to draw parallels here with the plateau argument first put forward by Bryan and Harter (1899). They argued that when performance at a task was plotted as a learning curve, one common feature of the curves was a 'plateau'. In many tasks, learning a new skill reaches a frustrating point of no apparent improvement in performance, though the task is



being practiced over and over again (learning curves illustrated in Appendix 6.). General performance appears to level off and may remain there, seemingly for ages, before acceleration occurs once more. Comparing this to the Haider & Frensch's (2002) data, learning the IR strategy may certainly lead to a jump or a discontinuity in performance. However, performance may not necessarily have flattened off before the 'discovery' of a reduction strategy.

Unlike the plateau argument, though, the main point here is not that individuals display a 'plateau', but that they may do so at different times along the curve from novice to expert. When the data from a number of participants are then averaged, the 'plateaus' are also averaged out. Under this proposal, strategy changes, of which IR may be one, do not violate the power law when the performance of individuals are averaged together.

Of course, taken a different way, this also indicates that using the power law as a benchmark of both skilled performance and competing theories may be misguided. The power-function may describe only special cases where discrete strategies do not occur. So discounting theories that do not follow such a continuous function may miss important realities in the development of task proficiency. Certainly IR, if accepted as a discontinuous strategic change, should not be discounted on these grounds.

### *Research questions*

A number of possible research questions are indicated by the work in this section. It would seem that most of what is known about IR comes from just one task. Consequently, there

may be attributes specific to this task that underlie the effect. This is important to test, particularly if IR is regarded as a general part of skilled performance.

A main issue is that IR was not stimulus-specific, as shown by the transfer across different stimulus-sets. This was particularly problematic for instance theories such as Logan's (1988). However, during transfer there was a rise in RT, suggesting that there was a memory component involved. A further investigation of this change is also warranted.

## **2.5 THE ROLE OF ATTENTION**

Information Reduction (IR) could be conceptualised as an attentional strategy. There is a strong attentional component required in its acquisition, task-redundant information is isolated and ignored, while task-relevant information is processed or attended. There are numerous, sometimes conflicting, models of focused attention, most of the debate centers around whether selection occurs early at a pre-attentive stage or later at a conceptual level.

### **2.5.1 EARLY SELECTION THEORIES**

The zoom-lens model (Eriksen & Yei-Yu Yeh, 1985, Eriksen & St.James, 1986) is an example of a weak early selection theory. The early selection view of attention follows the landmark research of Broadbent (1958), and proposes a two-process model of attention. At the first stage all the physical properties of a stimulus are processed in a parallel manner, this is followed by a second stage where more complex properties are processed (such as the meaning of words). This second stage processes in a serial fashion and so would be of more limited capacity than the first stage. To prevent an overload in the second stage, a

selective filter allows through only those stimuli that have a particular (or particular set of) physical property. This explains the well-known 'cocktail party' phenomena; it is possible in a noisy party to attend to the voice of one person while ignoring all the other voices. That is, the physical properties of the voice can be selected and all the others then filtered out.

Unfortunately, this theory will not explain all 'party' phenomena, for instance if someone else in the room mentions your name, your attention is sometimes captured and you start to listen to them. This suggests that not all distracters are filtered out. To account for this and other evidence that unattended stimuli are processed, Treisman (1960), proposed a less drastic version of Broadbent's filter theory. In her theory she proposed that unattended stimuli are 'attenuated' rather than filtered out. While usually the attenuation works so that unattended stimuli are not further processed, some stimuli have a pre-disposition to be relevant and thus even when attenuated, may still be strong enough to result in further processing. In the above example, your name would be very relevant and so would trigger processing even if you were not attending to the speaker.

### 2.5.2 LATE SELECTION THEORIES

The early-selection view, especially in its less drastic form has many proponents (Bundesen, 1990; Posner, 1980; Shiffrin & Schneider, 1977), however there is an alternate view. This is that all of the incoming stimuli, attended and unattended, are processed and selection takes place much later at a response stage. Still only relevant stimuli elicit a response and there is still a limited capacity 'bottle neck', but this is at the point of responding. To account for the fact that often very little is known about the unattended stimuli, it has to be accepted that this information is just forgotten very quickly. This would

seem to be a very wasteful procedure; however there is quite a body of evidence to support this view (Barnard, 1985; Deutsch & Deutsch, 1963).

### *Processing load*

Lavie (1995, 2000) suggests a plausible reason why this paradox seems to exist in the literature. Tasks that seem to support a late-selection theory tend to have low perceptual loads (e.g. with just a single target and single distractor, plus a relatively undemanding task for the target), while tasks that support an early-selection view have high perceptual load (e.g. more stimuli presented, and/or a more 'perceptually' demanding target task). This indicates that all the stimuli are processed only when there is enough resource available, or to put it another way, a limited capacity system is not exceeded.

Whether or not distractor information is processed deeply thus depends on whether the perceptual task for the relevant target exceeds this limited capacity. If the target task is undemanding, spare capacity may spill over to distractors and they will be processed to a greater extent. However, if the target task is higher in perceptual load, this may exhaust perceptual capacity, and so less distractor processing will take place.

To clarify the conception of perceptual load further, it is not simply equivalent to task difficulty. For example, tasks that are more difficult because of weaker stimuli, or because of greater memory loads, may not behave like those with higher perceptual load. Perceptual load is defined more as tasks which include more target stimuli and/or require more perceptual operations for the same target stimuli (Lavie, 2002). Cognitive load may also play an important part in attention. Unlike the perceptual mechanism above, which Lavie

feels is passive, cognitive load taps active cognitive functions concerned, for example, with working memory capacity and dual-task coordination (Lavie, Hirst, de Fockert, & Viding, 2004). Thus, perceptual and cognitive load may have dissociable effects on attention: the relative role of each type of task load are explored more fully in Chapter 5.3

The idea that the amount of information processed is dependent on load is also consistent with neuropsychological evidence that shows attenuation, rather than the elimination, of sensory responses to unattended stimuli (Driver, 2001). Attenuation could result from ‘winner takes all’ processes (a popular concept in visual psychophysics), in which neurons that respond to a stimulus are mutually inhibitory. Because not all the stimuli in a scene can drive all the neurons, a competition takes place. This competition is influenced both bottom-up by stimulus strength and top-down by current relevance. The ‘winners’ of the competition then attenuate or inhibit the ‘losers’. The more competition there is, the greater the level of attenuation. So the concept of attenuation, especially under high task load, seems supported from a number of converging sources.

### 2.5.3 SPACE-OBJECT BASED SELECTION

To some extent the early-late selection debate can be re-cast as one of space-object selection, so is attention driven by locations or objects? Again it is a question of how much processing takes place before attentional selection, is just a location specified or some higher grouping mechanism. The spot-light (Posner, 1980) or zoom-lens (Eriksen & St.James, 1986) metaphors are examples of a space-based attentional theory: from this perspective a particular region in space is ‘picked out or highlighted for further processing.

Treisman's Feature integration theory (Treisman & Gelade, 1980) is a special case of this type of model. In her theory the features of stimuli like colour and orientation are extracted in parallel pre-attentively just like the filter model, and then each feature location in space is attended to serially. In effect, it is attention to location that provides the 'glue' that binds together the features into the percept of an object. While there is a great deal of evidence to support the above models, there again are experimental situations that suggest attention may be directed to objects. Grouping processes certainly seem to effect performance if the stimuli are moving. McLeod et al (1988) found that if items moved together, they could form a perceptual group that aids target detection. For this to be possible, not only does attention have to be directed to an object, but such selection must be independent of the space that object occupies.

While there is confrontational dispute over space-based versus object-based models there is growing consensus (Bundesen, 1990; Driver, 2001; Lamy, 2001) that location is special, at least for visual attention. It seems intuitive to assume that to attend to something one must first locate it. It is possible, then, that even when a non-spatial property is attended, such as the colour red, it appears that this is done by attending to the location of that non-spatial property. It seems that even grouping processes exert their influence by directing the spatial distribution of attention (Driver, 2001).

The use of location may be the default in visual perception, only abandoned when a non-spatial property becomes more appropriate, such as for a moving object, where colour, shape or coherence may be more successful in directing attention (Lamy, 2001). To refer this back to the alphabet arithmetic task (AAT), it would appear that location was a vital element in information reduction (IR). When triplet position was varied, the strategy was

less successful, probably because attention had to be relocated constantly. But even then attention was directed by the location of the triplet.

The theoretical overview of attention suggests that while location may be important in the development of IR, this may not always be the case. Specifically, if the task involves moving stimuli or stimuli where non-spatial elements are more salient, movement cohesion or context may also drive IR. Again this is an empirical question. To more fully understand the properties of stimuli that may drive IR, different perceptual tasks that contain both task-relevant and task-redundant information will need to be realised.

### *Research questions*

The attention literature provides two further questions that may provide a basis for future research. These are;

1. Is there attenuated processing of unattended items.
2. Is this attenuation moderated by task difficulty.

One view of attention is that it serves as a 'filter' that attenuates unattended task components. There is some debate about this regarding early and late processing theories. Some synthesis of the two has been provided by Lavie (1995, 2000) who suggests that the stage of processing is dependent on processing load. In the case of IR then, it would be interesting to determine how much processing irrelevant task components receive and how task difficulty may moderate this aspect of task performance.

## **2.6 INFORMATION REDUCTION AND EXISTING SKILL ACQUISITION THEORIES.**

Haider and Frensch clearly consider their investigation into IR to be one that attempts to understand the development of skilled performance. As such, reduction in task processing needs to be accounted for in existing theories of skilled performance. However, what skill was becoming acquired in the AAT? This task does not readily come to mind when skilled performance is considered, and it does not really tap any motor skill. Haider and Frensch do though report a change in performance over trials with the AAT. Reaction times generally decreased, especially for the longer alphabetic strings. So, some cognitive or perceptual improvements were occurring.

It could be argued that it is not really a skill that was learned in the AAT, rather a rule or some compliance to task demands was acquired. It could similarly be argued that IR was not really different to the reduction in stimulus processing that could follow boredom with the task or a general lack of motivation. IR could also be considered less a skill and more a resource limitation on the part of the individual. For example a novice driver may know they need to look further ahead while driving, but lack the capacity to do this. The lack of resource in this case could reduce the amount of information attended. Under these arguments it does not appear that a skill is learned during experiments utilising the AAT.

However, it is unlikely that IR can be explained as a consequence of boredom, or a lack of resource; processing of the whole stimulus was not reduced, just the irrelevant items. In the driving example above, if resources were limited, relevant information would also be 'missed', leading to worse performance. IR does not lead to such a performance decrement,



as only irrelevant items are ignored, which speeds performance. The same applies to boredom or low motivation; participants were not processing the *whole* string at a lower level, which would lead to more errors over trials, they were rather increasingly ignoring just the redundant items. So IR does not appear to reflect a lack of resource or motivation, rather it reflects the development of a strategic choice not to process task irrelevant items.

The second argument is that it is not skill developing in the AAT, but rather just some form of rule abstraction. In other words, participants can abstract the rule that only the triplet items need to be scanned. However, in more general terms such a rule would specify where to look for relevant information during a task. This sounds much more like a skill. Take the driving example again. Processing all available information as you drive along would probably lead to overload and fatigue. Instead, the skilled driver knows where to focus attention dependent on context. At high speed, attention is focused further ahead, while in pedestrian areas the sides of the road near the car require more attention. There may of course be negative consequences of this information reduction strategy. Over time things that do not change may not be heeded because they are regarded as irrelevant, which is one reason why signs are put up to inform drivers that there is a 'new road layout ahead'. Without such a sign, changes in the road ahead may not be noticed quickly enough.

The AAT, then, is investigating a mechanism that may be an important part of general skilled performance. Participants make a strategic choice to ignore information in a task or situation they have learned is irrelevant to that task or situation. The AAT is a long way from providing a realistic demonstration of the generalisability of the mechanism though; this is why it will be important to demonstrate IR in a range of tasks in particular and

observe whether IR occurs in more naturalistic situations. In defense of the current approach, Lee and Anderson (2001) have discussed the decomposition of complex skills into simpler tasks. They conclude that simple tasks, as typically used in research, can scale-up to more complex tasks that are more typical of human learning in real-world contexts. Furthermore, they support the assumption that complex tasks consist of many smaller parts (in their case, the Kanfer–Ackerman Air-Traffic Controller Task), which are learned according to some basic learning principles. Thus, IR may be one important component of skilled performance which may be directly investigated using tasks like the AAT.

An attentional strategy does seem to occur in the AAT that serves to reduce the information processed. If this change over practice is regarded as one component of accomplished performance, it is a change that does not fit entirely with contemporary theories of skill learning. Most current theories of skill acquisition only explain qualitative changes in processing, that is, ‘how’ the information is processed. No account is given for situations where there is a change in ‘what’ information is processed. It is therefore important to determine how such quantitative changes can be incorporated into existing theories of skill acquisition.

#### 2.6.1 INSTANCE THEORY, ACT-R AND INFORMATION REDUCTION

Logan (Logan, 1988; Logan & Etherton, 1994; Logan, Taylor & Etherton, 1999) has expanded on his instance theory to include an ‘attention hypothesis’, this posits that attention determines what gets into an instance. Further, that an instance consists of co-occurrences and the effect of attention is to determine what co-occurrences are learned. In a

series of experiments Logan and Etherton (1994), found that the co-occurrence of word pairs was only learned to advantage when both words initially required attention. If attention was directed at only one word of the pair, any co-occurrences were not learned and so did not provide a performance advantage in their categorisation task. It is further suggested (Logan, Taylor & Etherton, 1997), that attention guides what is learned during automatization and they attempt to formalise a model that integrates Logan's instance theory with Bundesen's (1990) theory of visual attention. So it seems that attentional processes have a central role in the instance theory.

At first thought an attentional mechanism could integrate information reduction (IR) into the instance theory. If just the triplet becomes attended, then the co-occurrence of these figures will be encoded and eventually with practice immediately retrieved. This would certainly explain the performance improvement found during the AAT, even if the method of selection is not yet specified. Unfortunately, where this falls down is when transfer effects are considered. Haider and Frensch (1996) found that the performance gained following IR would also transfer to novel strings. Instance theory cannot account for this, as instances or episodes must be stimulus-specific. Perhaps one way that Instance theory could account for transfer would be if the concept of an instance became less constrained to include some generalised abstracted rule about information reduction. Logan himself acknowledges that 'similarity' must play a role in Instance theory. Certainly, when what constitutes an instance was considered earlier in the review (section 2.2.4), abstracting across instances seemed to be possible in a number of ways.

Anderson's (1998) ACT-R provides no better solution: it includes no mechanism by which there is a quantitative reduction in processing. Typically in this process-based account of skilled learning, performance is improved by speeding-up access to stored information, speeding the processing of information or both. This does not include any reduction in task processing. While at the rule level there may be some reduction in information as better productions are acquired, this does not change task processing so that task-redundant information could be completely ignored. Anderson (Anderson, Fincham & Douglass, 1997; Lee & Anderson, 2000) has more recently taken the view that skilled performance reflects a complex mixture of a number of processes.

## 2.6.2 A STAGE MODEL OF SKILL ACQUISITION

To this end Anderson et al (Anderson, Fincham & Douglass, 1997) have discussed a synthesis between ACT-R and Logan's instance theory, by arguing for a 'four stage model' of skill acquisition. In their experiments participants were presented with example statements such as *'Skydiving was practiced on Saturday at 5p.m. and Monday at 4p.m.'* The main point is that a simple rule may be abstracted, in this case the second skydiving event always happens **two** days later and one **hour** earlier (+2 -1). Participants received a number of examples like this, after which they had to complete missing time/day slots of further examples. Some of these were previous example statements while other were novel. The novel fragments would have to be completed by analogy with the learned examples.

From the results four different processes of skill acquisition were isolated. Participants start with analogy to examples, mapping learned exemplars onto novel statements to solve them. Eventually they become aware of a declarative rule (+2 -1), and apply this declarative

abstraction. With further practice participants develop production rules, which are the procedural embodiment of the rule. Finally, as the novel statements are repeated (they become novel no longer), straightforward retrieval of examples from memory solves the problem. These stages are not discrete and the method chosen depends on how well learned is that particular example, but it becomes clear that both an instance and a process account are being described.

Anderson et al (1997) suggest that the phase of declarative and procedural rule use corresponds to Logan's (1988) algorithmic method, while the retrieval from memory of learned examples aligns with Logan's example stage. From the point of view of the information reduction hypothesis, it is not the synthesis between productions and instances that are important, but rather that two different theories of skill acquisition are required to explain the data. Further, a process of analogy and declarative rule abstraction occurred to enable novel problems to be solved, so yet more processes are needed to account for transfer effects. It seems there may be no single answer as to how we acquire skills; rather a number of different processes may work in concert to provide performance gains.

#### *Further support of stage models*

Such a position is further sustained by Lee and Anderson (2000), they again propose a number of processes which underlie skill acquisition, one of which is information reduction. Using the Kanfer-Ackerman Air-Traffic Controller Task (KA-ATC), they found that improvements following practice were best described by three different categories of learning theory. First, they found that within the KA-ATC people changed the strategy they use to perform the task, such as not checking certain information. This they categorised as a

procedure selection model. Second, they suggest that participants strengthened their memories for the location of critical information. This they indicate is an instance of a strengthening model. Finally, they note that individual key-strokes became merged into macro-operators - sequences of key-strokes. This falls under the third category of a procedure transformation model. Lee and Anderson (2000) propose that contemporary models of skill acquisition all fall under these three categories, this again indicates that a single model will be insufficient to explain the mix of processes underlying the acquisition of complex task performance.

There is obviously a strong element of IR in the classifications above. Lee and Anderson, (2000) investigated this further by analyzing eye-fixation during the task. They split the ATC information screen into a number of areas, some of which they classified as task-relevant and some as task-redundant. Over practice they found the length of fixation on the task-redundant parts of the screen became significantly reduced. Improvement in performance at the task seems to rely in a large part on this strategy. This supports Haider and Frensch's (1996, 1999a, 1999b) earlier hypothesis that people can learn to distinguish task-relevant from task-redundant information and then learn to ignore the irrelevant information.

What is interesting though is that for the KA-ATC task, the areas of the screen classed as irrelevant were not completely ignored. While fixation for these areas were gradually reduced, the asymptote was non-zero. Two reasons are offered for this, participants may have had spare time as they got better and so searched less relevant information, or perhaps these areas were not as irrelevant as the experimenters hoped. This does suggest though that

IR is not the all or nothing shift implied by earlier studies (Haider & Frensch, 1996, 1999a, 1999b), IR may be a flexible strategy and attention may perhaps be distributed dependant on task. The flexibility is illustrated to some extent by the change in fixation on fuel level. In the KA-ATC task, planes queue and enter in a holding pattern, obviously they have to exit this and land before their fuel runs out or crash. However, as proficiency increases this becomes less of a problem, as planes are landed so quickly they never reach a critical fuel reserve. The result of this development is that participants gradually reduce fixation time for the area of the screen displaying fuel levels. This suggests that IR does not occur acutely, but rather is flexible enough to develop along with changes in competency, at least in the stated direction. The question is what can change IR in the other direction, so that attention is re-deployed to previously irrelevant areas if this becomes necessary?

### *Research questions*

There seems to be evidence that skill acquisition is a complex mix of processes that are only captured by some synthesis of existing models. However, the way in which IR and current theories of skill may integrate is still an empirical question requiring further research.

While there is some support that IR may generalise to more naturalistic tasks, yet more questions have been raised about its deployment. Particularly, it seems that irrelevant task components may still be processed to some extent even following the adoption of an information reduction strategy.

## **2.7 IMPLICIT LEARNING**

For IR to progress, the regularity of relevant and irrelevant information has to be noticed. That is, some mechanism that abstracts environmental regularity must be brought into play. In simple terms, how do people over practice notice which parts of a task or stimuli are unnecessary and may be ignored? Haider and Frensch (1999a, pp131) propose a two-step process, the first an implicit abstraction of the task regularity and the second a conscious decision to employ the IR strategy. If implicit processes, specifically implicit learning, play some role in IR, then some consideration of the area is prompted here. However, there is a very extensive literature concerned with this, so the review will be constrained to just sampling those reports directly relevant to the thesis.

Implicit learning as distinct from implicit memory is likely the most worthwhile avenue, since a working definition of the same bears directly upon IR. One way to define implicit learning is a process whereby knowledge is acquired incidentally about the structure of a stimulus or environmental regularity, which is then utilized without conscious awareness of having that knowledge (Kelly, Burton, Kato & Akamatsu, 2001; Mathews, 1995; Perruchet, Chambaron & Ferrel-Chapus, 2003; Reber, 1993). This definition suggests firstly, that knowledge is acquired without any intention or guidance and secondly, that this knowledge effects behaviour even if a person is unaware they have it. Subsequently such knowledge cannot be readily verbalised and has to be accessed or measured via other means such as preference, forced choice or guessing.



### 2.7.1 VERBAL EXPRESSION OF IMPLICIT LEARNING

Of more interest here is that such implicit knowledge may eventually become explicit. Further, the common view is that implicit learning mechanisms are capable of abstracting relevant from irrelevant features of a stimulus (Berry & Dienes, 1993; Mathews, 1995; Reber, 1993). This maps onto the two-stage process suggested by Haider and Frensch (1999a) and indicates a mechanism that can learn at a local level task relevant from task redundant information. In the first stage of the process, knowledge may be learned incidentally and without guidance, as found in the IR research. Additionally, in the second stage, this knowledge may eventually become explicit and so chosen as a conscious reduction strategy. In this second stage, it would be expected that the strategy could be verbally expressed.

Frensch, Haider, Runger, Neugebauer and Werg (2003) were particularly interested in the route from implicit learning to the explicit verbal expression of this strategic knowledge. They suggest that the abrupt performance discontinuities found in the AAT, described earlier in the review (Haider & Frensch, 2002), could represent the point when implicit knowledge became available to consciousness and could be verbalised. At this instant the knowledge was used to switch strategies from processing the whole stimulus to just the relevant information. As a result response times dropped sharply. The evidence was compelling and they indicated it was supported by other studies they have in preparation. Thus for their IR tasks, during practice a move was made through some implicit learning mechanism to implicit knowledge about the regularity in the task, this eventually become explicitly available immediately preceding a strategy switch to IR.

### 2.7.2 IMPLICITLY LEARNED ENVIRONMENTAL REGULARITIES

There are a number of papers that support the view that regularities in the environment may be learned implicitly in this way. Particularly relevant is a set of experiments using the McGeorge and Burton invariant learning task (Bright & Burton, 1994, 1998; Huddy & Burton, 2002; McGeorge and Burton, 1990; Stadler, Warren & Lesch, 2000). In the original version of this task (McGeorge & Burton, 1990), participants were exposed to a number of four digit strings each containing an invariant '3'. To divert attention away from this, a simple arithmetic calculation was carried out on the strings. Following this 'incidental' learning phase, participants were presented with pairs of novel strings and asked to decide which they had seen in the previous phase (though they had seen neither). The pairs contained one positive string with a '3' and one negative string without the invariant digit. They found that participants demonstrated a preference above chance for the positive strings.

This preference persisted even if the form of the strings was changed between learning and test phases, i.e. written numbers such as 'three'. Participants could not verbalise why they made these choices. McGeorge and Burton (1990), took this as evidence that an implicit rule was learned (each string contained the invariant '3') and learned this at an abstract level that allowed transfer to different stimulus structures. In a similar way one could suppose that a rule was learned about the regularity in the AAT strings; 'only the triplet is relevant to a response'. However, later papers (Cock, Berry & Gaffan, 1994; Stadler et al., 2000) indicate that such rule abstraction is unlikely; rather judgments are made on a similarity basis. This has led to an episodic account of implicit learning. However, at a

theoretical level, a task or environmental regularity still needs to be abstracted in some form from the episodes to allow such judgments.

### 2.7.3 'REAL-WORLD' IMPLICIT LEARNING

Since theories of skill acquisition should be applied to real world situations, some evidence that implicit learning happens outside the laboratory experiment would be encouraging. Unfortunately, this evidence is rather thin on the ground. Kelly et al. (2001) report that this may be due to poor test sensitivity; normally in experimental situations alternate forced choice methods are employed to 'get at' implicit knowledge. They consequently realised a 'real world' implicit learning paradigm using this forced-choice method and a number of common stimuli such as coins, stamps and beer labels. They found the regularities of these stimuli (direction the head faced) had been learned during daily life. This was demonstrated by a preference choice for the correct depiction, though importantly, this knowledge could not be verbalised. It would seem then that environmental regularities are learned in naturalistic situations, supporting the view that implicit processes are part of everyday behaviour.

There is some debate over implicit learning, even some question over the phenomena in general (Shanks & St. John, 1994). However whether rules are abstracted or episodic accounts are accepted, there does appear to be consensus that some mechanism exists that learns environmental regularities such as stimulus relevance.

### *Research questions*

As far as IR is concerned, some implicit mechanism seems a likely candidate to explain how relevant information is noticed in a task from irrelevant information, even if this knowledge does eventually become available to report explicitly. This requires further research in order to determine the exact nature of the mechanism that leads to IR.

## **2.8 FUTURE DIRECTIONS**

The review substantiates the argument that one element of skill acquisition may be the deployment of an attentional strategy that serves to reduce the quantity of information processed. This is achieved by a two-step process, first task-relevant information is segmented from task-redundant information, and then task-redundant information is ignored. This has been termed information reduction (IR) by Haider and Frensch (1996, 1999a, 1999b), and provides a challenge to existing theories that capture only qualitative processing changes during practice.

### *Redundant information reduction*

While it will be necessary to integrate quantitative changes in general into existing theories, IR itself is not yet fully understood. An initial and important question is by what means do people segment task-relevant from task-redundant information; how is the deployment of attentional resource decided? Green and Wright (2003) use a modified version of the AAT, to provide a possible answer. In their study they make the distinction between task-irrelevant and task-redundant information. In the Haider and Frensch (1996, 1999a, 1999b) experiments the post triplet part of strings was task-irrelevant during the practice blocks.

The latter part of the strings were irrelevant to task completion, in other words the post-triplet string contained no useful information. Green and Wright's (2003) question was this, would IR occur if strings included task-relevant information that was duplicated by other task-relevant information within the strings. In other words, if the task could be carried out by attending to either or both of two task-relevant information sources would reduction still occur. Would task-relevant but duplicated information be ignored.

Green and Wright (2003, Experiment 1.) found that when there were two co-occurring errors in the practice strings, IR still took place. This was evidenced by a reduction in the string length effect and a following increase in post-triplet errors during the test block. In the test block errors were manipulated to no longer reliably co-occur, so attending only to the triplet lead to mistakes. Overall, this suggests that information that is task-relevant, but redundant due to the duplication of prediction, can also be ignored. So the IR effect still occurs when there are two task-relevant sources for string verification. However, if both sources, triplet and post-triplet sequences, are equally valid, how is the choice made to attend to one while ignoring the other? One possibility is that information is selected by the order in which it is processed

#### *Time of processing.*

In a follow-up experiment (Green and Wright, 2003, Experiment 2) the time each source was processed was manipulated. This was done either by moving the triplet to the end of the string (E F G [4] L), or by requiring participants to read the string in reverse – from right to left (L K J I [4] D). In both the triplet-last and reverse-order conditions the IR effect was again found for correct strings. Analysis of the test block errors revealed that for the

triplet-last condition, participants detected more pre-triplet than triplet errors, suggesting they were not using the triplet as their primary information source. For the reverse-order condition a reversal was found, now participants detected more triplet than post-triplet errors. Suggesting in this case, that the first encountered source became the primary source of information.

This reversal in error types seems to indicate that the information source first processed, becomes chosen and relied upon. Green and Wright (2003) conclude from this, that the time of processing is a factor in the information source chosen when a task contains more than one source of task-relevant information. They propose that people come to recognize the co-occurrence of the two sources and then chose to attend to the first processed. So time of processing may influence the operation of the IR mechanism.

As Green and Wright (2003) note, there are a few difficulties in completely accepting this hypothesis. For instance, the two information sources are not entirely comparable, for the triplet-last condition, the first encountered source (E F G in the example above) is just a recognition task of an over-learned alphabetic sequence. This is a much easier computation than the triplet-source (G [4] L). Whereas, the same cannot be said of the reverse-order condition, in this case the post-triplet segment has to be processed in the reverse to the normal order and so is not a simple recognition task. So in this latter instance, triplet and post-triplet processing load may be more similar. As Green and Wright (2003) suggest, it would be interesting to determine the effect of differing processing demands on which source is chosen. If IR operates as a form of cognitive economy as indicated in the paper, then the source that imposes least demand on resources should be preferred.

Green and Wright (2003) proffer a mechanism of IR where the co-occurrence of reliable information sources are learned and time of processing may lead to preferential choice amongst these sources. However, it is possible in their study that participants do not notice/learn the correspondence of triplet to post-triplet errors. For example, if the error is in the triplet and feedback shows them to be answering correctly, they may never learn to look at the post-triplet error or realise its co-occurrence. So, the basic idea remains the same, in that first processed information gets priority, but the assumption is lost that a co-occurrence between sources is learned and then one source is ignored. Another area for future research, again mentioned by Green and Wright (2003), focuses on the reliability of each source in providing task accuracy. In the AAT above, each source was completely reliable and consistent, in real world situations this may not always be the case.

### *Research questions*

The reliability of stimuli in providing accurate information could be another factor affecting how participants choose to reduce to one source? As mentioned earlier, the relative task difficulty of redundant and relevant task components may also moderate reduction? These are two interesting questions for further research.

## **2.9 THESIS RESEARCH QUESTIONS**

The literature review both of skill acquisition in general, and the work of Haider and Frensch (1996, 1999a, 1999b, 2002) in particular, has helped to identify a sequence of research questions. To some degree these are interlinked.

### 2.9.1. GENERALISING INFORMATION REDUCTION

To regard information reduction (IR) as a general part of skilled performance requires that it occurs in many tasks that have some element of redundancy. To date most of what is known about IR comes from one task, the AAT. It is possible that this task may have specific attributes that make reduction particularly favorable. An important question, then, is whether IR will generalise to other tasks that contain irrelevant information? This will only become clear as other paradigms that promote IR are realised.

### 2.9.2. STIMULUS-SPECIFIC COMPONENTS

Due to the constraints of the alphabet only a limited set of unique strings were possible in the AAT, this may have lead to some set-specific learning taking place. While Haider and Frensch (1996) found IR to persist across changes in stimulus-sets, RT did increase after the set switch, suggesting some item-specific learning had been lost. In a later study, Haider and Frensch (2002) acknowledge this, but do not investigate it further.

#### *Mnemonics*

Initial testing with the AAT suggested that it is quite easy to form mnemonics to aid triplet recall. For example if the string was H [4] M N O, then H [4] M could be remembered as referring to 'him'. This means the 'skip four letters' computation would no longer be necessary, speeding verification times even more. It is possible that with repetition, H [4] M is just recognised as correct. It becomes obvious that this is describing an 'instance' (Logan, 1988) account of learning. Participants move from a computation to learning instances with the aid of mnemonics and then possibly end up performing by direct retrieval of the solution from memory. This suggests that two strategies may occur in the



AAT, one reduces the information processed while the other encodes instances of the attended subset.

One avenue of research, then, could determine whether IR still takes place when stimulus strings are less easy to learn and appear not to repeat. Theoretically, an important question relevant not only to understanding the mechanism of IR, but also the place of the information reduction hypothesis in contemporary accounts of skill, such as Logan's (1981) instance theory and Anderson's (1993, 1998) ACT\*.

### 2.9.3 TASK ATTRIBUTES

There is also question of the stimulus or task attributes that lead to IR. Further intrinsic task properties such as overall task difficulty, segmentation of the information sources and repetition of task instances would give more insight to the antecedents of an IR strategy. Reviewing the literature of attention and strategy use suggest these could be fruitful areas of investigation. The work of Green and Wright (2003) further suggests attributes such as task difficulty and the regularity of any redundancy should be investigated. From a practical perspective, if IR is a common property of everyday skilled performance, understanding the conditions likely to promote and moderate reduction is important. This is particularly important considering previously that just one particular task, the AAT, has been repeatedly used.

In the attention literature, location seems to hold a special place for focused visual processing and it appears that in both the AAT and KA-ATC, information reduction progresses by selecting locations to attend and ignore. While location may be the default

attribute to guide both attention and IR, it is possible other attributes may be more appropriate for other tasks. Such as tasks that have moving targets, in which case, colour, shape or motion coherence may 'guide' IR, rather than just location. Again from a practical viewpoint, this would be important to understand for dynamic displays such as radar and ATC operations.

#### 2.9.4 INDIVIDUAL DIFFERENCES

In addition to the difference between tasks, there are also the differences between individuals to consider. Haider and Frensch (1996, 1999a, 1999b) report that not all the participants in their studies 'reduce', some carry on processing all the available information throughout the whole session. As noted by Green and Wright (2003), these non-reducers may provide another avenue of investigation. This difference could be due to personality type or the level of anxiety, whichever, there are distinct individual differences present in the performance of the AAT.

#### 2.9.5 IRRELEVANT TASK COMPONENTS

The fate of task-irrelevant and task-redundant information also calls for further study. Even though there is evidence in this review that task-relevant information becomes fixated, the attentional field may spread beyond this. Additionally, Lee and Anderson (2000) found that participants in the ATC task often fixated on task-redundant information, especially if they had time.

From a practical perspective this is an important issue, while some element of a task may be redundant or even irrelevant at one point, at another it may prove vital. In the KA-ATC

task participants learned that fuel-level was redundant information and so ignored it. It takes little imagination to picture a scenario where the fuel-level of planes could suddenly become very important, and then a serious problem could easily develop. This is a rather extreme case, but illustrates the point: it is important to understand how much redundant information is processed and how this unattended element may recapture attention. How flexible is the return from IR?

#### **2.9.6 CONSCIOUS AWARENESS OF THE STRATEGY.**

In addition to these main areas for research, there seems little agreement about participant's awareness of the IR strategy. Lee and Anderson (2000) report that participants were not really aware of the strategy they were applying following practice. On the other hand Haider and Frensch (1999a) found IR to be affected by instruction, suggesting that it was under conscious control. However, in the same paper they describe a two-step process; the first step, isolating task-relevant from redundant information may be an implicit process. While the second step, the choice to ignore redundant information may be a conscious one. It is also possible that the level of awareness may begin as implicit, but move to explicit knowledge. After reviewing the literature this is another possible avenue of research.

### **2.10 SUMMARY AND INITIAL FOCUS**

This literature review was wide ranging, since skilled performance taps many cognitive and perceptual abilities. Previous studies of IR have also taken this view, drawing on a wide range of domains and literature. This is understandable at this point, since the way IR fits with contemporary theories is yet to be determined. Similarly, the task properties that promote and moderate IR are only beginning to be understood.

The review of the literature highlights one possible problem with the current research in IR; most of what is known comes from the alphabet arithmetic task (AAT). If IR is a general part of skilled performance, one aim of the thesis could be to generalise the effect to different paradigms. A second string to the research could investigate the intrinsic task conditions that promote and moderate IR. To some extent this is linked to the previous goal, since it requires a new task for comparison. Often when task strategies are reported, a switch to a new strategy is suggested. This implies a new method is adopted suddenly and without return. Another further question then, is can IR be regarded as a strategy 'switch'; one that results in completely ignoring the unattended stimuli? The attention literature seems to suggest otherwise, so this is an empirical question to pursue.

As Haider and Frensch (1999a, 1999b) explained, paradigms to investigate IR are difficult to operationalise. This may be why they chose to increase the scope of their results in successive experiments by producing variants on the same alphabet arithmetic task (AAT). For the same reason this thesis also initially focuses on a variation of the AAT. The main aim in running the first study was to attempt a replication of the basic IR effect. To extend the results a little, it was also decided to investigate the consequence of adding 'noise' to the stimulus. That is, would IR generalise to more naturalistic noisy stimulus situations? The first study used the same strings as Haider and Frensch (1996, Experiment 2), but with the letter-case randomly varied across strings to add the aforementioned 'noise'. This is especially interesting since it may also make the triplets more difficult to learn over practice. Depending on the results of the study, the next step may take a number of directions.

If mnemonics play a part in task performance, at least for the AAT, the IR effect in the initial study could be reduced compared to earlier results. This may suggest the change in letter-case disrupts this memory component. The next logical step would employ a task where items are not repeated, so negating any 'instance' learning account. Following this, some investigation of the degree of processing directed at task-redundant information would be interesting and would build on earlier work by Haider and Frensch (1999b). These initial findings should then provide grounding from which to progress to other possible attributes that lead to IR.

## **CHAPTER 3:**

### **OVERALL METHOD**

*This section explains the types of task used in both previous information reduction (IR) research and also this thesis, along with the analytic methods used. Additionally, the different task manipulations are summarised without embedding them in theory or results. Finally, any additional issues with the general method are listed along with a small scale study designed to determine how much IR may be ascribed to general practice effects.*

### **3.1 THE ALPHABET ARITHMETIC TASK (AAT)**

Haider and Frensch (1996) proposed that over practice any redundancy in a task may be isolated and ignored. To investigate this information reduction (IR) hypothesis they employed an ingenious version of the alphabet arithmetic task (AAT) and also developed the following analytic method. As described earlier the task consisted of verifying whether strings of letters followed the alphabet or not. What made the task more difficult was that a calculation was also necessary, since the strings contained a bracketed digit. This digit indicated the number of letters to skip in the alphabet during the verification. So for example F[4]KLM is a correct string, since 'K' follows 'F' when four letters in the alphabet are skipped, and the letters following the 'K' are alphabetical. Thus the string can be verified positively as alphabetic. However F[4]LMN is an incorrect string since five letters are skipped between 'F' and 'L', even though the letters following the 'L' are alphabetical.

Crucially within the task, for incorrect strings, not only was the incorrect letter displaced by one position in the alphabet, the incorrect item was always immediately after the brackets. So the error position was always regular and part of the calculation involving the digit. This letter-number-letter segment of the string will be referred to as the 'triplet' from now on. In the standard task (Haider & Frensch, 1996) this triplet could be followed by 0-4 extra post-triplet letters, so string length could vary between 3 and 7 items in total. Strings began with letters 'D' through 'M', this avoided using the very start or end of the alphabet. So strings starting with 'D' for correct strings were:-

D [4] I or D [4] I J or D [4] I J K or D [4] I J K L or D [4] I J K L M, for lengths 3 to 7 respectively. Similarly, incorrect strings were realised as:-

D [4] J or D [4] J K or D [4] J K L or D [4] J K L M or D [4] J K L M N,

So for each starting letter (D, E, F, G, H, I, J, K, L, M,) there were five possible correct and five incorrect strings, so in every block of the experiment there were 50 correct and 50 incorrect strings. From 3 to 30 of these *training* blocks were used in previous experiments; with strings following the above format. As can be seen from the examples, any error occurs directly following the brackets, which relegated the letters following the 'triplet' as redundant for the verification of each string. Thus there was an underlying regularity built into the strings, this can be expressed as the rule that 'only the triplet needs to be processed'.

### 3.1.1 MEASURING INFORMATION REDUCTION WITHIN THE AAT.

The rationale was that initially the whole string would be processed. However, if the regularity was detected, the post-triplet may eventually become ignored as trials progressed. Since the strings varied in length, to begin with reaction time (RT) should increase with the number of items. If reduction occurred, this should however change over practice as the extra post-triplet items are not longer processed. As a result RT will no longer be tied to the length of a string.

The change in RT brought about by the number of string items will be termed the string-length effect throughout the remainder of the thesis. One important thing to point out is that the string-length effect will only appear for the correct strings, as the post-triplet items are searched for an error. For the incorrect strings, since this is a two-alternate forced choice paradigm, as soon as the triplet error is detected a response should be made. Therefore RT should be similar across all incorrect string lengths. So it is the correct strings that were



most interesting as far as IR was concerned, since this string type would indicate any reduction in the string-length effect.

### 3.1.2 DETERMINING THE STRING-LENGTH EFFECT WITH REGRESSION ANALYSIS.

To determine whether the string-length effect does reduce over practice, following Haider and Frensch (1996), a regression model with string-length as the predictor and RT the dependant variable was used. In this model while string-length was a good predictor of RT regression coefficients should be high. So at the beginning of a session regression slopes should be steep. If though reduction occurred, these slopes should flatten since RT would be relatively less affected by an increase in string items. To test this statistically, for each participant and each block the best fitting linear regression lines across the five string lengths were computed. The result was subjected to ANOVA where the within-subjects independent variable was block and the dependent measure was the regression coefficients.

If the slopes declined over blocks significantly this directly indicated a reduction in the string-length effect and indirectly suggests IR was developing over practice. Indirectly, because there are reasons other than IR that could account for the decline. The alternative argument is that the post-triplet items were processed faster and faster as the blocks progressed. This would also lead to a reduced string-length effect over time, explaining the change in mean coefficients. Another alternative is to propose that reduction did occur, but this was not a change in strategy, rather just a fatigue effect. As fatigue set-in over the many trials, just less and less attention was given towards the ends of the string. This is particularly convincing since it is normal to read from left to right. To counter this latter

criticism Haider and Frensch (1999b) ran an experiment where the triplet was at the end of the string rather than the beginning. They still found a reduction over blocks somewhat ruling out the fatigue explanation.

### 3.1.3 ERROR RATE FOLLOWING IR.

To further exclude the possibility that the post-triplet was not ignored, but rather processed ever more quickly over practice, a different final block was used (Haider & Frensch, 1996, 1999b). Typically this 'test' phase followed the training blocks with no indication to the participant. In this final phase the regularity of alphabetic errors was changed. For a percentage of the incorrect strings alphabetic errors were moved to the post-triplet string segment. The principle was that if the redundancy was actually ignored, these post-triplet errors would be missed. Alternatively, if the post-triplet items were just processed very quickly the errors should be caught and the appropriate response initiated. This formed the second thread of Haider and Frensch's criteria measuring IR. Thus, if regression coefficients declined over blocks and post-triplet test errors were missed, this converging evidence confirmed IR. That is, a strategy to reduce the information processed develops within the paradigm following task experience (Haider and Frensch, 1996).

## 3.2 THE TARGET SEARCH TASK (TST)

Everything known about IR to date (Green & Wright, 2003, Haider & Frensch, 1996, 1999a, 1999b, 2002; Haider, Frensch & Joram, in press) comes mainly from studies that use variations of the AAT. Subsequently, it was thought necessary to develop a different task for the thesis to both compare and generalise the original findings. The development

and rationale for this task are laid out in the appropriate chapters, but there is a description in this section for completeness and also to render understandable the following test of practice effects.

The new task was called a target search task (TST) and methodologically was very similar to the AAT. Strings of letters were verified as to whether a target was present or not. Each string varied in length from 3 to 7 items and each string was generated to conform to an underlying regularity or rule. This was that if a target was present, it was consistently at the second position from left. A memory set of three targets were given and these were searched for in each string, only one target per string would appear. Thus, the underlying principle of both tasks was very similar, if participants reduced, then they would ignore items beyond the second position since they were redundant for string verification. Conceptually though, the two tasks are quite different, since the TST requires no calculation, there is no perceptually different segment and so the task has one goal, to holistically search the whole string. The strings were for example of the form:-

S V G K W - for a five item target present trial and

F B Z A E I N - for a seven item target absent trial.

In this case the memory set was C,O,V.

Subsequently the analysis for this task followed that set out by Haider and Frensch (1996). If reduction to the relevant information occurred then varying the length of the string would have gradually less effect on RT. Additionally, if targets appeared during a final test phase in the formally redundant segment – a novel or untrained position - they would be missed

leading to an increase in errors for this target type. This method was used for three of the thesis experiments and each was analysed in an identical way to that described for the AAT.

### **3.3 SUMMARY OF TASK VARIABLES – IN THE LITERATURE**

#### *length of practice*

The AAT has been used a number of times under various manipulations (Green & Wright, 2003, Haider & Frensch, 1996, 1999a, 1999b, 2002 and in press). A short summary of the main points of the methods used is included next. The intention is to highlight the various manipulations rather than the theoretical reasons or implications for using them.

IR has been investigated at 1, 3, 5, and 7 blocks (Haider & Frensch, 1996, Expt. 2), but also in as many as 10 blocks, (Haider & Frensch, 2002) or even 30 blocks (Frensch et al., 2003). Typically, each block has 100 trials. Overall, IR seems to have developed by 500 trials or 5 blocks.

#### *Rule abstraction*

Apart from the typical paradigm set out above the first new manipulation attempted by Haider and Frensch (1996, Expt. 3) was to split the strings into two stimulus sets. One starting with D, E, F, G or H, the other with I, J, K, L or M. One or other of the sets was given first for four blocks then switched for the other set for a further four blocks. IR persisted across this switch in stimulus sets. Thus IR occurred without instruction and this self-discovered rule was not stimulus-specific.

### *Task instruction*

The next variation was for a study Haider and Frensch ran in 1999, using different task instructions. The instructions were to work either as quickly as possible or as accurately as possible. In Experiment 1 the conditions were, referring to the instructions given, speed, accuracy, speed then accuracy, accuracy then speed and a control (be as quick and as accurate as possible). 8 blocks were run; for mixed conditions the instructions changed after 4 blocks. So for example participants were told to be as quick as possible for the initial four blocks, then told to be as accurate as possible for the remaining four blocks. In this case each block contained only 30 correct and 30 incorrect trials per block. This was because string lengths were only 3, 5 & 7 items. Feedback was given at the end of each block.

In Experiment 2 in this paper, to try and eliminate the difference between the speed condition and the speed-instructions part of the accuracy-speed conditions, additional time pressure was added to speed-instruction phases. The time pressure was added to encourage Ps. out of their previous high-accuracy strategy. For the 'speed' condition or phase, there was 2000ms presentation for block 1, minus 100 ms per block thereafter. For the 'accuracy-speed' condition, there was 6000ms presentation during accuracy part then presentation duration matched the speed condition for the speeded phase. For the other pure 'accuracy' condition, presentation time was 6000ms throughout. IR was moderated by task instructions, increasing under speeded conditions.

### *Triplet position and eye tracking*

Haider and Frensch (1999b, Expt. 1) in their next paper manipulated triplet position. There were three conditions; triplet first, triplet last and triplet at random positions during

training. 50 correct and 50 incorrect strings as usual and for the test trials 20 strings with errors in one of the four possible novel locations.. Interestingly only half of the triplet-first/last groups received feedback for test trails when error was in a novel position. There were 8 training blocks and string length was the usual 3-7 items. IR developed for all conditions, though the effect reduced for the mixed condition.

In the second study reported in this paper, eye tracking was used to directly measure reduction. Only 0, 2 or 4 post-triplet letters were used this time (strings of 3, 5 & 7 items). Strings were 30 correct and 30 incorrect during training, so 8 blocks of 60 strings were created. Conditions were fixed triplet first and fixed triplet last as the previous experiment. Note that the strings were very big, 3.8cm letters 6 cms apart. Feedback was only at the end of a block, this time for all the blocks, training and test. The new dependent variables were mean number and mean duration of fixations on the relevant and redundant string positions. It appears that only the triplet was fixated following reduction.

#### *Redundant vs. irrelevant information*

Green and Wright (2003) took a new approach and manipulated the relevance and redundancy of string items. Basically, they created strings so that if the triplet was correct the post-triplet was also correct and the reverse, if one was incorrect so was the other. So information was not so much irrelevant to verification as redundant. The direction the triplets were read, left-right or right-left was also varied by instructing the participants on how to scan the strings. The conclusion was that *redundant* information was also ignored and the source processed first was relied upon. This briefly summarises the manipulations

realised in the literature, with the theoretical considerations stripped away. Of note is that they all used variations of the AAT.

### **3.4 SUMMARY OF TASK VARIABLES – IN THE THESIS**

Similarly, in the thesis a number of manipulations have been tried, but these have used both the AAT and the TST.

#### *Letter-case randomisation*

The first study in the thesis used the AAT in the same fashion as Haider and Frensch (1996, Expt. 2) except this time the letter case was varied randomly from trial to trial. That is, the letters in the strings were a random mix of either upper or lower case. There were 6 training and 1 test block. IR still developed in this task.

#### *TST and string length*

Study two used a new task, the TST, in which a memory set of targets were searched for amongst a string. In the first of two experiments the strings were the same lengths as for the AAT – 3 to 7 items long all uppercase. There were two training conditions, one training block and then a test block, or five training and a test block. In a second experiment, the maximum string length was increased to 13 items, thus the lengths 5, 7, 9, 11 & 13 were used with 10 examples of each. IR was found in both experiments of the study, but was perhaps more convincing in the second experiment.

### *Intrinsic task attributes*

Study three in the thesis investigated the contribution of intrinsic task attributes to IR. The experiment in this study varied three task attributes. These were string segmentation, task difficulty and string familiarity. For the first of these brackets were placed around the first three items from the left. For the second, an alphabet calculation was required that compared the target, if present, with adjacent letters. The aim was to determine if either came before the target in the alphabet. The third attribute was varied by reducing the number of unique strings per block, so in effect they were recognisable as repeating. There was also a standard condition that used the strings from study 2 for comparison. Compared to this standard, there was some moderation of IR when the various attributes mentioned were added. As before there were five training and one test block

### *Regularity of error position*

The final fourth study of the thesis manipulated the regularity of item redundancy. In 100%, 90% and 60% conditions alphabetic errors were manipulated to occur never outside the triplet, 10% outside the triplet and 40% outside the triplet respectively. Feedback was given at the end of each of the five training and test blocks. Some slight speed instructions were given; to work quickly. The regularity of redundancy was found to moderate IR.

This briefly summarises the manipulations attempted so far in the tasks used to investigate IR. Overall the effect appears to be robust, developing under varied conditions and in two different tasks.



### **3.5 DEBRIEFING QUESTIONNAIRES**

#### *Rationale*

One approach to understanding the development of IR would be to ask participants to verbalise what they are doing as they progress through the task. This type of protocol analysis though is extremely time consuming to record, later code and analyse. In reality this could form a whole other line of research, so was not attempted here. Instead a short end of session questionnaire was administered, as some report from the participants was thought useful. The questionnaire was kept purposely short for reasons of time as expressed later. The focus of the questions was to draw out what participants knew about the underlying regularity within the task. That is, did they feel their approach to the task differed as they went along? Also, whether they noticed the targets appeared in a consistent position.

The questions were created so that first they were very open and then asked explicitly if the regularity was noticed. Typical examples of the questionnaires are in appendix 2. One caveat to point out here is that when explicitly prompted about the error or target position, participant report is probably subject to experimenter or demand effects. That is the question then infers something about the strings which the participant is more likely to agree than disagree with. This often happens with for example eye witness testimony (Lindsay, 1990). So results from this type of question have to be viewed with this in mind.

### **3.6 OTHER ISSUES WITH THE METHODS USED**

#### *Non-reducers*

On reading the above descriptions of the tasks used to investigate IR, it may seem that discovering the strings contain irrelevant or redundant items is trivial. After so many trials where a target did not appear at certain locations, developing strategies to take advantage of this regularity could be expected? However, in each study ran for the thesis, there were always a number of participants who did not notice that part of the string could be ignored. This was also reported by Haider and Frensch (2002) and Haider, Frensch and Joram (in press) for their studies. In which they report a figure of around 40% were non-reducers for the standard task.

To some extent this leads to ‘noisy’ data, but reduction does still happen consistently enough for IR to robustly replicate across a number of studies, which increases confidence that it is a real effect. More importantly, it also indicates that noticing the regularity of target placement and using this information is neither easy or inevitable. The reason why some individuals do not reduce is not yet known, but should be the subject of research.

#### *String-length and error feedback*

There are a number of other issues worth pointing out here, particularly with the work of Haider and Frensch (1996, 1999a, 1999b). They seem to have taken away feedback in some of the studies, while IR still occurred the effect of this is not fully addressed. Also they switch from 5 string lengths (3,4,5,6 & 7 items) to 3 (3,5 & 7 items), with no real explanation as to why. It could be supposed that such a move would make the task more

sensitive to reduction, since it was the intermediary lengths that are removed. Possibly this would make the effect of string-length on RT more 'dramatic' and subsequently more likely to decline with reduction. They also suggest that for incorrect strings reduction will not occur, as the search is stopped on encountering the error. In separate analysis they find exactly this, with the slopes running along close to zero, since only the triplet is processed in every block for this string type. So one could regard this as a baseline; what coefficients become under full reduction. It would seem interesting then to analyse both correct and incorrect string regressions together, this does not seem to happen in the literature but was used in the thesis.

### 3.6.1 EXPERIENCES OF RUNNING THE TASK

#### *Task effort*

On a more general note, experience with running the studies for the thesis has highlighted some further issues. The first is that the AAT and to a lesser extent the TST are both arduous repetitive tasks. Blocks of 100 trials require continued concentration and some level of self-discipline on the part of the participant. This was unavoidable as the investigation involved measuring changes in performance over practice. The rationale was to discover if individuals learned what part of a task required attention and which part could be ignored as irrelevant. If this is a general part of skilled performance, it should be learned over practice at a task. To this end, the participants were required to practice a novel task until the expected improvements in strategy occurred. Nonetheless, this did mean the experimental sessions were effortful, repetitive and boring.

This was one of the reasons why, for the thesis, a maximum of six blocks were administered, which took approximately one hour to complete. Even at this level a number of participants voiced relief that the task was over at the de-briefing. A secondary consideration was that the usual student participant panel did not exist at the Open University because it is a distance learning institution. As a result, the few post-grad students and mainly staff on site were approached to participate. Mostly, participation involved lunch hours or time off work, so the length of a session had to be reasonable and around an hour. This was not thought problematical as the evidence points to IR occurring by 500 trials or five blocks (Haider & Frensch, 1999a, pp131).

#### *Participant availability*

The lack of any participant panel and participants available only during lunch time meant running studies was very time consuming, with just one or two run per day, so the experimental part of the thesis took some time to complete. A large amount of time was spent recruiting and chasing up participants from various sources and eventually an attempt was made to form a list of possible volunteers by sending out a site wide email and posters around the site. Unfortunately due to the uninformed nature of the task – participants had to be unaware of the underlying task regularity - once a participant had done an IR task they could not do so again. Consequently, the list of willing volunteers had to be used and de-briefed very carefully, except for the last study this meant using small group numbers. A number of other universities were approached, to ask if studies could be run with their students, but none felt this was feasible. Except that is for the final study where part of the Brunel student panel was offered. Fortunately, IR appeared very robust and was found with

cell numbers below those used by Haider and Frensch (1996, 1999a, 1999b), which is encouraging for this line of research.

### *Confounding strategies*

Turning to some other confounding variables, particularly within the AAT, there appear to be the formation of strategies within strategies. Of particular note was that some participants reported using mnemonics to aid memorising the triplets. The most obvious reason for this was to avoid the repetitive triplet calculation. Though this does suggest that for the AAT there was some part of the task that was seen as repeating and could be learned. This is something Haider and Frensch (2002) themselves point out, also indicating participants learned the triplets. This is quite interesting as IR is set against and outside the usual instance account of skill acquisition.

One other variable to be monitored when running these studies was the exact form of instructions used. The paper on task instruction (Haider & Frensch, 1999a) reports instructions for speed and accuracy affect IR. To a lesser extent other forms of instruction to participants may also moderate the effect. For this reason, the thesis study instructions were either read from a script or appeared 'on screen'. The form of instruction was kept as constant as possible except for the last study, when some speeded direction was introduced.

### 3.6.2 REGRESSION COEFFICIENTS

One of the indicators of Information Reduction (IR) used initially by Haider and Frensch, (1996) and also throughout this thesis, is an analysis of the any reduction in regression

coefficients over practice at a task. The coefficients are the result of modelling the number of string items or string-length as a predictor variable against the response variable of reaction time (RT). The rationale is that if task redundant items later in a string become ignored as practice progresses, string-length will predict RT to a lesser and lesser degree and coefficients will become smaller.

Haider and Frensch in all their published work focusing on IR have used the unstandardised regression coefficients or  $b$  in their analysis (Frensch, Haider, Runger, Neugebauer, Voigt & Werg, 2003; Haider & Frensch, 1996, 1999a, 1999b, 2002; Haider, Frensch & Joram, in press). This thesis has followed their example to make the work here directly comparable. However, one could question if the standardised regression coefficient or beta would be more appropriate. Both  $b$  and beta are routinely within the output of current statistical packages.

Within, particularly, multiple regression one would wish to determine the relative contribution of each predictor variable. One way to ascertain this is to note the relative size of the regression coefficients between each pair of variables. However,  $b$  may be a poor measure of this if the *predictor* variables were measured on different scales. The unstandardised coefficients are highly influenced by each variable's mean and standard deviation (Lehman, 1995). Thus, any relative difference in these may effect the degree of slope, irrespective of how well that variable predicts the dependent measure.

One way to partially overcome this problem is to use the standardised regression coefficients or beta. This transforms the set of coefficients replacing  $b$  with  $b$  times

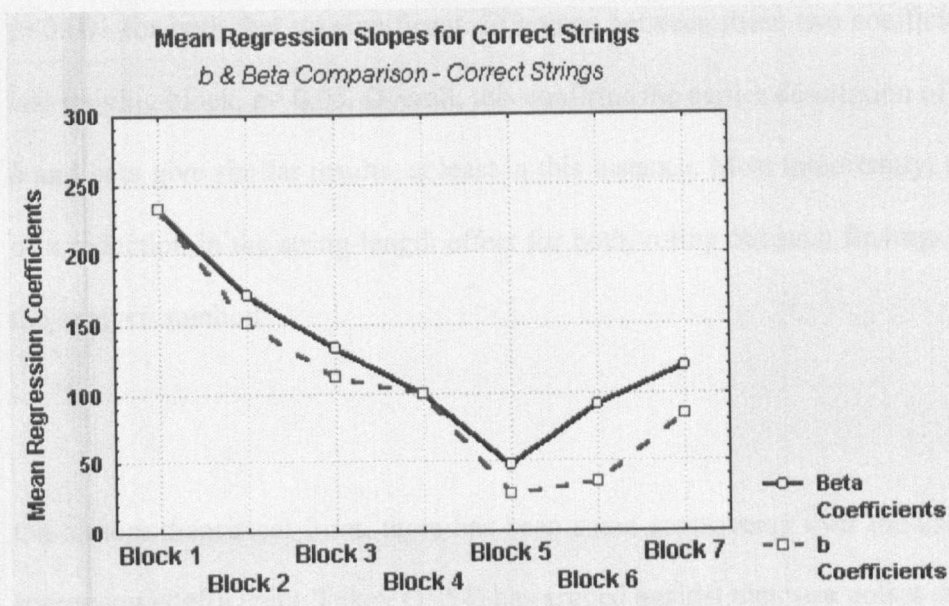
$s(X)/s(Y)$  where  $s(Y)$  is the standard deviation of the dependent variable,  $Y$ , and  $s(X)$  is the standard deviation of the predictor,  $X$ . This gives *all* the variables a mean of 0 and a standard deviation of 1. In effect, reducing them to *z scores*. The particular advantage of course is that the predictor variables become equivalent no matter what the original scale of measurement. The relative contribution of each variable within a model is still influenced by the presence of the others though, so even beta can not be taken as an absolute indication of a variables influence.

How does this relate to IR and whether  $b$  or beta should be used? While linear, but not multiple regression was used in the IR analysis, in actuality, the size of the regression coefficients across blocks was being compared to determine if reduction was taking place. To some extent one could argue that no problem exists with this as the predictor variable was always the same; the string-length. As suggested above beta is normally chosen when the predictor variables are of different scales, which is not the case here.

One possible problem though, was that as RT decreased over blocks this in itself may have reduced the value of  $b$  irrespective of any change in the string-length effect. That is, the reduction in slope of blocks found using  $b$  was really an artefact of the overall reduction in RT as practice progressed. There is little doubt that general speed-up and IR are confounded in these analysis. However, attempts to remedy this by using beta as an alternative may also be problematic because of this link and so may reduce the sensitivity of the analysis to IR.

One straightforward way to examine this is to plot the regression using both  $b$  and  $\beta$  and compare them. This was done for the data from the first study in the thesis (4.2.2) using the AAT. Of main concern was to determine if IR was still indicated by declining slopes when both  $b$  and  $\beta$  were plotted. To this end the data illustrated are for the correct strings as these are where IR would have an effect.

$\beta$  was less than 1, so to make the  $\beta$  and  $b$  values comparable all the  $\beta$  coefficients generated were multiplied by a constant. The results are illustrated in figure 3.1 below.



*Figure 3.1: Mean regression coefficients (b & beta) from Study 1, AAT, 7 block condition.*

As can be seen, the function provided by the coefficients for standardised and raw scores is very similar, importantly both show a reduction in value between the first and last training block. This suggests that the string-length was having less effect as training progressed.



These results were subjected to 2-WAY ANOVA (Coefficient Type -  $b$  or beta) \* (Block – 1 to 7) as independent variables and the coefficient values as the dependent variable.

There was no significant main effect of Coefficient type  $F(1,10) = 1.45$ ,  $p > 0.05$  (as expected since beta was multiplied by a constant to make it equivalent to  $b$ ). There was a significant main effect of Block  $F(6,60) = 3.79$ ,  $p < 0.01$ . However, there was no reliable Coefficient Type \* Block interaction  $F(6,60) = 1.03$ ,  $p > 0.05$ . Post-hoc tests (Tukey) revealed a significant difference between first and last training blocks for both  $b$  & beta,  $p < 0.001$  for both, but no significant difference between these two coefficient types for the last training block,  $p > 0.05$ . Overall, this confirms the earlier description of figure 3.1; both  $b$  and beta give similar results, at least in this instance. Most importantly, there is evidence of a reduction in the string-length effect for both, ruling out such findings as an artefact of the analysis method.

On a more theoretical front, there has been some controversy over the use of standardised regression coefficients. Tukey (1954) has argued against their use unless the dependent and independent variables have the same variance. Which in the case of the IR data they do not. Blalock (1971) follows this view, suggesting that standardised coefficients obscure the casual law and complicate the sampling distribution. Kim and Ferree (1981) have forcefully argued against the use of standardised coefficients, suggesting selected appropriate variables should be standardised beforehand and then the un-standardised coefficients used.

Perhaps a more serious criticism as far the analysis of the IR data is concerned comes from Marsden (1981, p16). He points out that the formula for standardising coefficients utilises group-specific measures of the predictor and dependent variables in transforming the original un-standardised coefficients  $b$  into standardised coefficients. The problem occurs when two or more groups are compared as the group-specific variability may not be the same in the different groups. This is a problem for this type of comparative research.

To explain this in a way more specific to the analyses in this thesis, suppose that variances change as the experiment goes on. Since beta (standardized coefficients) is related to both the  $b$  and the ratio of dependent and independent variable variances, a change in the latter might increase or decrease the beta, even if  $b$  were unchanged, or might even mask a change in the beta. Therefore, if the variance of RT changes over blocks this might influence beta irrespective of the change caused by string-length. If the main criticism of  $b$ , as used in the IR analysis, is that it may be open to some systematic change in RT, then beta maybe no better.

Some other points to note are that when  $b$  is used to determine IR, the target present slopes do not decline in the same way as the target absent slopes, this suggests indirectly that the un-standardised coefficients are not reducing over blocks just because RT is changing overall. So, it would seem that the method is sensitive to a reduction in the string-length effect. Additionally, in the data here and in the literature, the slopes seem to decline more when test errors for targets outside the trained position increase. Again this indicates that the  $b$  regression slopes used are sensitive to changes in IR.

In conclusion, it is difficult to determine whether the raw or transformed metric would be the most appropriate to use from the AAT data presented above. However there is some controversy over the use of beta. Certainly the standardised coefficients may not be comparable over blocks if changes in the block-specific variability changes. Again, then, there is no clear reason to move away from the measured metric and use transformed data. The method as developed by Haider and Frensch (1996) certainly seems sensitive to changes in information reduction, and the effect has been replicated a number of times. Taking these points into consideration the use of the un-standardised  $b$  coefficients was retained for the thesis, especially since this will make the results directly comparable with published work in the field of IR

### 3.6.3 INDIRECT MEASURES OF IR

Regarding the measurement of IR, in the typical AAT or TST paradigms, IR is really only measured indirectly. The degree of reduction can only be inferred from the regression and error analysis. In an attempt to improve this situation Haider and Frensch (1999b) ran an experiment using eye tracking equipment. Using this method the eye fixations of the participants could be determined. While this seems a valid experiment and suggests that following IR it is the triplet that is mostly fixated, there is one major flaw with the study. To make eye tracking feasible with their equipment they made the strings very big, 3.8cm letters 6 cms apart. This rather changes the task since scanning such a large stimulus is likely to be more effortful than the usual size which was 0.3cms high and 0.2 cms apart. It could be argued therefore that reduction may not be entirely at a perceptual level with these

shorter strings, since the span of attention may include some or all of the redundant items. This raises the possibility that IR may also occur partly at a cognitive level.

Accepting this, all the measures of IR in the thesis were indirect, so the levels of processing must be continually questioned. Particularly any alternative explanations for the analysis results should be interrogated. One such alternate explanation is that the general results rather than illustrating a reduction strategy are instead just a measure of fatigue effects. Also it is more than likely that there are general practice effects in the studies, people generally get faster with practice and this could explain at least the reduction in the string-length effect over time. That is, the extra string items would have less effect with practice. A short study follows where these considerations of fatigue and practice are explored.

### **3.7 PRACTICE EFFECTS – WHEN NO IR IS POSSIBLE**

#### **3.7.1 RATIONALE**

One criticism levelled at the IR hypothesis was that it explained just a fatigue effect (Haider & Frensch, 1999b). That is, people were not actively reducing the information they processed, rather as they become fatigued they were less inclined to process the later items in a string. For both the AAT and TST this could explain the IR results, as the affect of both string-length and error (or target) position could be changed by fatigue. Alternately, the string-length effect could also reduce over trials because participants just became quicker at processing the strings. In this case the extra items will have less and less effect as they are scanned ever quicker.

Klapp, Boches, Trabert, and Logan, (1991) using a version of the AAT found this sort of general practice effect. They report that following practice, extra addends to the string have a reduced effect, particularly when a level of automisation was reached. Indicating that for alphabet arithmetic tasks in particular, performance does improve just with practice. However, in the reported experiments (Haider & Frensch, 1996) error data suggests quicker processing does not explain the results. Quicker processing would not explain why targets in novel positions were missed during the test phase. Logically, for this to take place the post-triplet part of the string must necessarily have been ignored.

On the other hand, attenuated attention could still provide an alternate explanation to IR. Loss of attention or effort while scanning the strings could explain both the reduction in the string-length effect and why during the test block participants missed errors (or Targets) when they appeared in novel untrained positions. First, paying less attention to the longer strings for later blocks will reduce the length effect in a similar way to IR. Second, at test, targets appeared in an untrained position at the latter part of the strings. So as participants became fatigued towards the end of the experiment they could understandably miss these. Fatigue could thus explain both the reduction in regression coefficients over blocks and the increase in missed target when they appeared later in a string.

To counter this argument, Haider and Frensch (1999b) ran a study where the triplet appeared at the end of a string rather than the start. They still found IR and so proposed that it was not just a fatigue effect. Though, IR was much reduced in a second condition if the positioning of the triplet was randomly first or last in a string. Green and Wright (2003) also used a condition in which the triplet was placed at the end of a string. Again they

found IR. If fatigue in scanning from left to right, as usual for reading, was a complete explanation of the results in the above instances, the triplet should be eventually ignored when it was at the end of a string. This did not happen, so this straightforward view of fatigue does not explain away IR. However, placing the triplet at the end of a string does not entirely rule out fatigue as an account. Participants could still focus on the triplet and so could just gradually put less and less effort into the rest of the stimulus as they tire.

In fact, Green and Wright (2003) found that when instructed to do so, participants could scan the strings in the opposite direction; right to left. So it is possible that the triplet may still be processed first, even when it is at the end of a string, after which fatigue reduces processing of the remaining items. Of course, as tasks get monotonous or stressful this could promote a search for an effort reducing strategy such as IR, so the two probably are linked to some extent. Accepting this, there is likely to be some element of fatigue explaining IR test results, the question is, is it all a fatigue effect? Also there are likely to be general practice effects as trials progress, processing will get quicker once the task is fully understood and the responses become mapped to the stimuli. Again the question is, is it all practice effect?

The answer to the latter point is probably no, IR does not just reflect practice effects such as quicker processing of the strings or faster responses, the error data at test do not support this as already explained. Additionally, the participant self-report data in this thesis suggest that some individuals (between approximately 40%-90% across studies) were aware of the consistency of error (or target) positioning, suggesting it is not 'just' fatigue causing the data pattern. It would still be interesting though to determine how much of the change over

practice was not due to IR. Perhaps a better test of this would be to run the task so that all the blocks are similar to the test block. For the TST this would mean that the target now appears throughout the string, so reducing would lead to an increase in errors. Under these conditions it would be expected that the string-length effect for target absent trials would not reduce over blocks, and errors would not increase for test targets. That is, those further right than position 2 in a string during the last block. As indicated there is already substantial evidence that IR is more than just practice or fatigue, so just a short study was run to further add to this.

### 3.7.2 METHOD

Everything was kept constant from thesis study 2 (see chapter 5); the only change was that every block now consisted of the same 'test type' block as used elsewhere. 25 of the target present trials were constructed so the target appeared at second from left in a string. The other 25 of the target present trials were constructed so the memory item appeared in the string further right than second from left. So could appear in position 3 to 13 from left. The position was counterbalanced across trial as far as possible. The order of string presentation was randomised across blocks and participants. Seven subjects were run in this version of the experiment, ages ranged from 16 to 48.

### 3.7.3 RESULTS

The two results of interest were how the regression slopes declined over blocks and also the pattern of errors\* target position for the last block. The data was analysed in the same fashion as set out above and followed the criteria set out by Haider and Frensch (1996)

TARGET ABSENT/PRESENT STRINGS: STRING-LENGTH EFFECT DURING ALL BLOCKS.

As can be seen from figure 3.1 the regression coefficients start out higher for target absent than target present trials as still expected. Even though the target now appears throughout the string, this is at random and so not tied to the length of the string. Hence string-length is still not a good predictor of RT for target present trials. In comparison string-length is a better predictor of RT for target absent trials; participants are searching the whole string. The slope does however seem to reduce slightly over trials for target absent strings.

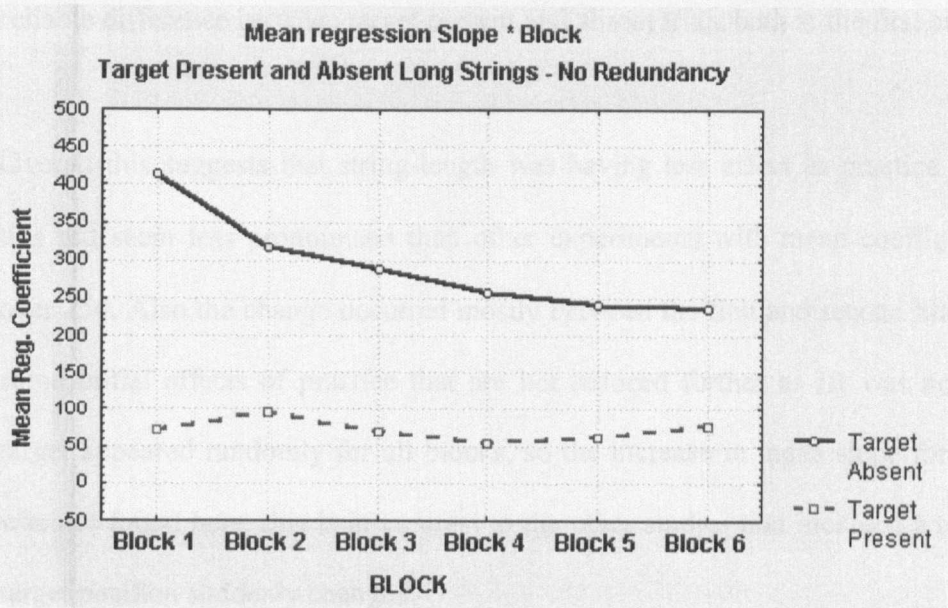


Figure 3.2: Mean regression coefficients over all blocks for Target Present and Target Absent strings; Practice Experiment.

When the coefficients for the target present and absent were analysed a 2-Way ANOVA Target(Present, Absent – within subjects) \* Block(1 to 6 – within subjects) revealed a



significant main effect of Target  $F(1,6)= 14.59$ ,  $p<0.001$ , and a main effect of Block  $F(5,30)= 6.24$ ,  $p<0.001$  and a Target \* Block interaction  $F(5,30)= 5.44$ ,  $p = 0.01$ .

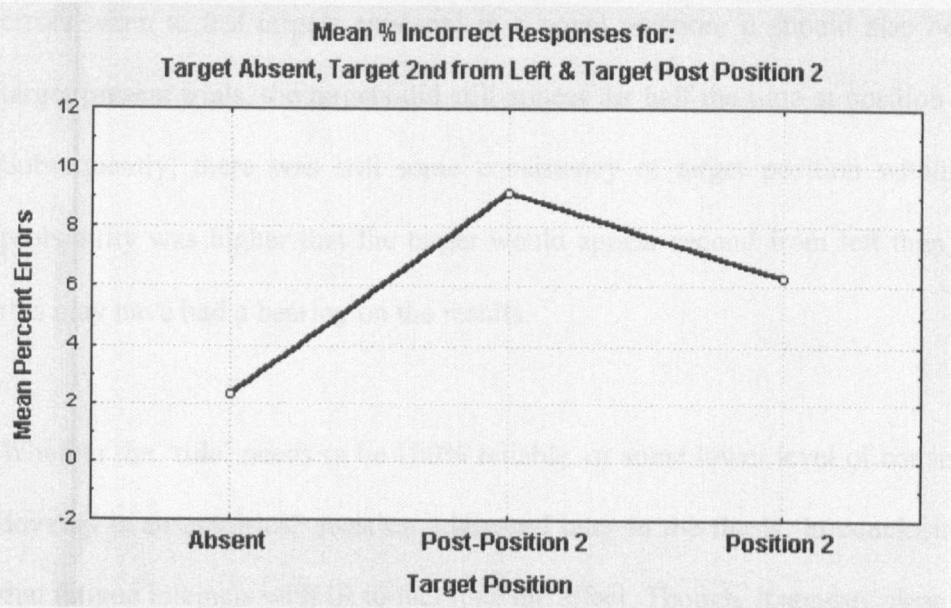
Post-hoc (Tukey) tests revealed that for target absent trials there was a significant difference between the first and last block ( $p<0.05$ ), though this difference came mostly from the change between block 1 & 2. As revealed by post-hoc(Tukey) tests again, for target absent trials, block 1 was significantly different from block 2 ( $p<0.05$ ), but block 2 was not reliably different from any of the following blocks, 3, 4, 5 & 6 ( $p> 0.1$  in each case). There was no reliable change across blocks for target present trials. There was a reliable difference between target present and absent trials both at the first and final blocks.

Overall this suggests that string-length was having less effect as practice progressed, but this did seem less pronounced than other experiments with mean coefficients remaining over 250. Also the change occurred mostly between the first and second blocks, suggesting some initial effects of practice that are not reduced further as IR was not possible. The target appeared randomly for all blocks, so the increase in mean slope for the final block was not found here, this is in contrast to the other studies that included a test block where target position suddenly changed.

## ERROR RATES

As can be seen from figure 3.2, there were more errors when a target was present than absent and slightly more errors when the target appeared beyond position 2.

A one-way ANOVA 3(Position – Target absent, at position 2 or post position 2: within subjects) did not quite reach a significant main effect of target position ( $p= 0.059$ ), any effect coming most likely from the target present vs. absent difference. A follow-up One-Way ANOVA, Position(Target at position 2 or post-position 2: within subjects) to test this assumption, still found no reliable difference ( $p>0.1$ ). This indicates that the position of the targets did not affect error rates. So it appears that the whole string was processed throughout this study, rather ruling out a fatigue explanation of the other results here and in the literature.



*Figure 3.3: Mean percent error rate for the final block where target could appear 2<sup>nd</sup> position from left in the string, or randomly in a position beyond this; target absent string errors added for comparison; Practice Experiment.*

### 3.7.4 DISCUSSION

The regression data did suggest that slopes declined, especially from the first to second block, the error data suggested high error rates for the last block though this was not affected by target position. Overall, it does seem that some practice or fatigue effects were present in the TST: particularly it seems the string-length effect can decline over blocks without IR. However, this does not explain the full pattern of behaviour found in the later TST studies, which must still be attributed to IR. The much more acute drop in regression slope over blocks and the reversal at test in this later study strongly indicated something additional to fatigue occurred. This was further supported by the significant increase in errors when at test targets appeared in a novel position. It should also be noted that for target present trials, the targets did still appear for half the time at position two in a string. Subsequently, there was still some consistency of target position within the task. The probability was higher that the target would appear second from left than anywhere else, this may have had a bearing on the results.

Whether the 'rule' needs to be 100% reliable, or some lower level of consistency for IR to develop is an empirical question addressed later in the thesis. In conclusion it does seem that fatigue interacts with IR to facilitate the effect. Though, it appears clear that individuals do learn to actively ignore consistently redundant items and that this leads to increased errors only when this consistency changes. It seems that human information processing is complex within such tasks, there was likely to be more than one process at work at any one time, and these may even interact (Anderson et al., 1997; Lee and Anderson, 2001).

## **CHAPTER 4: STUDY 1**

### **Information Reduction in the Alphabet Arithmetic Task: Random Letter Case**

*Within this chapter an initial experimental investigation of the information reduction effect was conducted. The letter-case of the string stimuli were randomised to determine the role played by stimulus familiarity during practice.*

## 4.1 INTRODUCTION

In their second experiment Haider and Frensch (1996) found that with practice irrelevant information within a task becomes unattended, a change they termed Information Reduction (IR). Further, they found that participants could discover redundant elements in the task without guidance from the instructor concerning this redundancy. This finding has significant implications for the understanding of skill acquisition, in that practice not only changes *how* information is processed, following the contemporary view, but possibly also *what* is processed.

Contemporary theories of skill acquisition and automaticity suggest that through practice, there are changes in the way that information is processed (Anderson, 1998; Fitts, 1962; Logan, 1999; Newell, 1990; Shiffrin and Schneider, 1977). For example Logan's (1999) instance theory suggests a shift from an algorithm method to the retrieval of solution to a problem from stored instances. Similarly, Anderson's (1998) Act-R theory posits a change from declarative to procedural rules over practice. Neither theory though accounts for a change in the amount of information processed when skilled performance at a task is achieved. Therefore it would seem important to understand IR, especially considering the working memory limitations and cognitive economy of the human information processing system.

Considering the work of Haider and Frensch (1996, 1999a, 1999b), it seems intuitive that information reduction strategies could become available when real-world tasks are engaged and may partly account for improved performance following practice. The generality of the

phenomenon is open to question, as all the experimentation carried out so far by Haider and Frensch investigating IR has utilised only one task; the Alphabet Arithmetic task (AAT) as described earlier. They have chosen to investigate the effect by manipulating this one task in favour of developing different paradigms.

The reason for this, as they suggested (Haider & Frensch 1999b p.173), was that generating a new task was far from trivial. Such a task would need to be novel to participants; so the effects of practice can be mapped without the variation caused by prior experience. Also the task has to be mastered and allow improvement over a relatively short experimental session; especially if a general pool of non-specialist participants is to be used. The task also has to contain irrelevant information that is both noticeable to the participants and may be ignored without response error. In addition to all this, the paradigm would have to allow the measurement of IR as it occurs. Producing such a task equivalent to the AAT in these factors is difficult, so the alternative route of extending findings by manipulating the AAT was attractive. The obvious shortfall of this strategy was that some of the properties of IR so far discovered may be particular to the AAT.

One problem with the AAT was that it closely followed a reductionist approach; very controlled presentation of the stimuli, which always followed a very consistent form. However, in contrast, naturalistic tasks while consistent, to a degree usually have some variation from one instance to another. For example lighting conditions may change, or objects may be viewed from different angles. So, stimuli outside of an experimental setting are inherently 'noisy'. Therefore, in day-to-day tasks, the IR effect would have to be robust enough to occur even when such noise is present. This may be a particularly important

criticism when the exact make-up of the alphabet task is scrutinised. While it is true that the task contained one hundred different strings per block, this was only true if the whole string is considered. The combination of the letter-number-letter-triplet and post-triplet addends meant 50 correct and 50 incorrect unique instances could be created. If however, just the triplet is focused upon, the number of novel instances drops dramatically. In fact in Haider and Frensch's experiments there were often only ten correct and ten incorrect triplets. This was a number that could at least be recognised as repeating if not rote memorised.

Considering the difficulty in the triplet calculation and the number of times it had to be completed, it would not be surprising if alternate strategies to avoid the calculation were sought. Some experience with the task indicates that it was possible to form mnemonics out of the letter pairs that make up the triplets. For example H[4]M makes HAM and made this triplet easy to recognise as correct when it repeated, thus avoiding the computation. So during training certain triplets may have 'flagged' whether a string was correct or incorrect and these would then always lead to a correct response. At least until the test block, when the rule change occurred and sometimes letters did not follow the alphabet post-triplet.

There seemed to be some evidence that learning of the triplets occurred, particularly if Haider and Frensch's (1996 Expt 3) transfer study is examined. In this study the set of strings were divided into two sub-sets, participants trained on one sub-set and were then transferred to the other. They found that the string-length effect was only slightly affected by the letter-set change, suggesting that the mechanism for IR is largely stimulus-independent. However, while IR persisted when the stimulus sets were changed, there was an increase in RT at the switch. The explanation given for this was that participants were

confused when the letter-set changed. Whilst this is possible, they were not so confused that they started to scan the whole string again. So perhaps a more likely explanation for the RT increase is that they were no longer able to recognise the strings or triplets and so had to revert to the more time consuming triplet calculation; increasing RT. While this experiment supports the view that instance retrieval is not necessary for IR to persist, some consistency or familiarity of the strings may still be necessary in its development. Certainly, familiarity plays a role in implicit learning and IR may initially progress at the implicit level (Frensch, Haider & Runger, 2002)

The ability to learn to recognise over practice that strings are repeating, or to even generate mnemonics of the triplets to aid such recognition, seems possible. The question is whether this attribute, particular to the AAT, played a role in the formation of IR? It does seem intuitive that the consistency of the strings would affect the level of confidence to ignore their irrelevant elements. From the perspective of a participant, a string was presented that has the same triplet as seen a number of times previously. This may be recognised and so the post-triplet part of the string was also unlikely to change and so can be ignored.

Schunn (1997) found just such an effect of stimulus ‘familiarity’ on strategy choice. Following earlier work (Miner & Reder, 1994; Nelson & Narens, 1990; Reder 1982, 1987, 1988), Schunn (1997) found that a ‘feeling of knowing’ underlies strategy choice, specifically the choice of whether to retrieve a solution from memory or compute it. Reder and Ritter (1992) suggest this feeling of knowing was driven by familiarity with the problem, in particular, depending on how similar the problem was to those seen previously.



Whether IR can be viewed as a strategy is an empirical question; certainly from the work of Haider and Frensch (1999a) it appears to be a conscious goal driven preference. As such it falls under the accepted definition of a strategy as ‘any procedure that is non-obligatory and directed toward a goal’ (Newton & Maxwell, 2000; Seigler and Jenkins, 1989). However, if IR has some level of automaticity in its deployment, it may ‘run’ without requiring conscious decision. This would make reduction obligatory in some sense, and move it outside the definition of a strategy above. However, accepting that IR can to some extent be employed as a strategy, the evidence so far may indicate that string recognition is important for IR to develop within the AAT. This may be critical in understanding both the mechanism involved and the conditions under which IR is likely to develop. One way to investigate this would be to introduce more ‘variation’ into the strings across trials.

## **4.2 EXPERIMENT 1: LETTER-CASE-ALTERNATION**

Introducing more variation into the strings is restricted due to the number of letters available in the alphabet. Excluding the first few over learned letters means there is a strict limit on the number of unique strings that can be generated. One possible alternative that allows the introduction of additional variation within the strings is to change the letters visually while retaining the identity of the letter. One way to achieve this is by randomising letter-case over trials, such as ‘E [4] j K l’ and then ‘e [4] J k L m’. There is a great deal of literature investigating lexical access that employs this type of letter-case alternation, and seems to suggest that changes in case can affect letter recognition (Cole & Haber, 1980; Posner, 1967; Posner & Keele, 1967; Posner and Taylor, 1969).

Posner *et al* (1969) have suggested that in a same-different letter naming task, responses are longer when the case is alternated. This, they suggest, results from visually presented letters simultaneously being retained in both a visual code lasting under two seconds and also a 'name' code that lasts several seconds. When the case is alternated, the faster physical recognition has to be eschewed for the slower name code. Avant and Thieman (1985), again using a same-different paradigm, found that letter-case, orthographic regularity and lexical/semantic information are analysed in an automatic fashion. Overall this suggests that the alternating letter-case in the AAT is a change likely to be processed and that matching repeating strings across trials may become more difficult, at least at a visual level.

If word recognition is considered, Mayall, Humphreys & Olson (1997) report that case mixing can disrupt performance at many tasks, such as, letter identification, word and non-word naming, lexical decision, semantic categorisation and syntactic categorisation. In all these tasks, case alternation affected performance compared to when the case remained constant. Though, the size of the disruption appears to depend on the exact stimuli used (Mayall *et al* 1997). They suggest that the source of this disruption is the inappropriate grouping of letters within a word, which makes retrieval of the word from the lexicon problematic. One could imagine that case-alternation may also make the triplet instances more difficult to retrieve.

Overall, it does seem that case-alternation is likely to disrupt recognition performance at the visual level for the AAT strings and this in turn may make the stimuli seem less familiar for string repeats across trials and blocks. Support for this last point comes from a number of studies that use case-alternation specifically to reduce the configural features of words in

word-recognition experiments (Besner, 1983; Lavidor, Ellis & Pansky, 2002). This manipulation seems to disrupt performance for words but not non-words, and is taken as evidence that case-mixing disrupts both early letter coding and a familiarity check mechanism. Words are familiar, and this familiarity is affected by case-alternation, non-words are not familiar so are not subject to the same disruption. It is possible this familiarity check mechanism may also come into play for the non-word AAT strings as they are learned and become familiar over trials.

Besner (1983), describes such a familiarity check or familiarity discrimination mechanism (FDM), that uses configural features to make certain kinds of decisions about words. This mechanism may be similar to the hippocampal FDM described in the animal learning literature (Besner, 1983). It seems that the FDM is a mechanism that signals familiar patterns overall and specifically may be used as a source to bias decisions in lexical tasks. It is possible then, that this FDM process may also play some role in signaling the familiarity of string patterns over trials. Accepting that strategy choice is moderated by familiarity (Schunn, 1997), it is likely that this aspect of the AAT strings may be disrupted by case-mixing. Not only will case-alternation make the task a little more naturalistic by introducing ‘normal’ stimulus variability to the trials, the strings should also appear less consistent and familiar as the triplet region repeats within and across blocks.

The aims of running the first study therefore, were two fold. First, to replicate the earlier results of Haider & Frensch’s second experiment (1996); determining that our methodology captures IR and our analyses are appropriate. Second, to ascertain whether IR still occurs after the introduction of stimulus ‘noise’ in the form of random letter-case across trials,

such as 'E [4] j K l' and then 'e [4] J k L'. If IR still occurs with the modified stimuli, confidence is increased that it may be part of the performance gain found as tasks are practiced in every day situations.

#### **4.2.1 METHOD**

##### **DESIGN**

A three factor mixed design was used – Test Block Error Type(Triplet vs. Post-Triplet – within subjects) \* String Length ( five levels, 3-7 – within subjects) \* Blocks (3 or 7 – between subjects). The dependent measures were mean error rates per block and mean RT per block.

##### **PARTICIPANTS**

24 participants were used, randomly ascribed to either the 3 block or the 7 block condition. Participants were employees and students of the Open University Milton Keynes and their ages ranged between 20 and 52 with a mean age of 34.8 years. For participation each was paid £5 per hour, so dependent on condition (3 or 7 blocks), participants received either £5 or £10

##### **STIMULI/MATERIALS**

Following Haider & Frensch (1996), Participants verified either 3 or 7 blocks of training alphabetic strings followed by one test block. Each training block contained 50 correct and 50 incorrect strings.

For the incorrect training strings, an alphabetic error (alphabetic error is used to describe the part of the string that does not follow the alphabet, this will distinguish it from response errors which the participant makes) occurred within the letter-number-letter triplet, following the number [4] without exception. This regularity built into the task relegated the post-triplet string as redundant to correct task performance. The number '4' was always used to make the experiment comparable to the original. The alphabetic strings were constructed in the same way as the original experiment, a letter-digit-letter triplet followed by 0,1,2,3 or 4 further letters. However, in this present study the case of each string letter was randomly varied between uppercase and lowercase. To take care of the randomisation, a small Visual Basic program was written that randomly generated a string of 1's and 2's (1= upper case, 2 = lower case). Presentation of the trials was randomised between participants and blocks. The experiment was generated and run using the Experiment Run Time System (ERTS, BeriSoft, 1998)

In addition to the practice blocks two further sets of strings were generated. One set of ten practice strings, used to ensure participants fully understood the task and one '100 string' test block. The test block contained 50 correct strings, 30 incorrect strings with the error in the triplet and 20 incorrect with the error outside of the triplet (the part of the string that was redundant during the practice stage). Location of the post-triplet error in these later strings was counterbalanced so that it occurred in various string positions. The strings were in Times New Roman 18 font size yellow script, presented on a mid-grey background. The screen, viewed from a normal distance, was of 1024\*768 resolution. Letters were approximately 4-7mm in width and height, spaced 3-4 mm apart, each dependent on case.

This meant string-length varied between 26 mm to 60mm for 3 to 7 item strings respectively.

## PROCEDURE

The experimental session begin with each participant reading an informed consent form (appendices) that gave details of the experimental task and took the place of an initial briefing, any questions that arose were answered at this stage. If in agreement with the content of the form it was signed and the experiment commenced. Participants received computerised instruction about the task, explaining how strings were verified as correct or incorrect, followed by examples of in/correct strings. Importantly no mention was made about the regularity of alphabetic error position.

After this, the ten practice strings were presented. A central fixation cross appeared for 500ms, replaced immediately by the alpha-numeric string, which remained until a response was made. Participant responded using the 'Z' or 'M' marked key on a standard keyboard, for example 'Z' for a correct string and 'M' for an incorrect string (keys counterbalanced for half of the participants). Following the response, participants received feedback consisting of a correct message or error message with simultaneous tone. The message was on screen for 500 ms and then the screen cleared for 1000ms before the next fixation cross.

Any participant with more than 3 errors during practice (i.e. over 70% errors) went through the instruction and practice phase again. Any participant failing to reach the criterion at the second attempt continued on with the experiment, but their data was excluded from analysis

(they were run through the complete experiment on ethical grounds, so that they did not feel they were rejected because they were not capable of the task). Following this practice phase, participants were presented with either three or seven training blocks. The stimuli appeared in an identical fashion to the practice strings; each block took approximately 15 minutes and participants were allowed to rest as they wished between blocks.

After the last training block and without any further instruction, the test block immediately followed, in this block errors occurred in the formerly redundant post-triplet part of the string. The test block was the final set of strings to verify and marked the end of the experiment, participants were thanked for their participation and given a short questionnaire to determine how they undertook the task and if they noticed the regularity built into the training trials. Finally, a brief description of the rationale for the study was given and then any questions answered, after which the participant received payment and was invited to leave. The 3 training block condition took on average 63 minutes to complete and the 7 training block condition took an average of 97 minutes to complete.

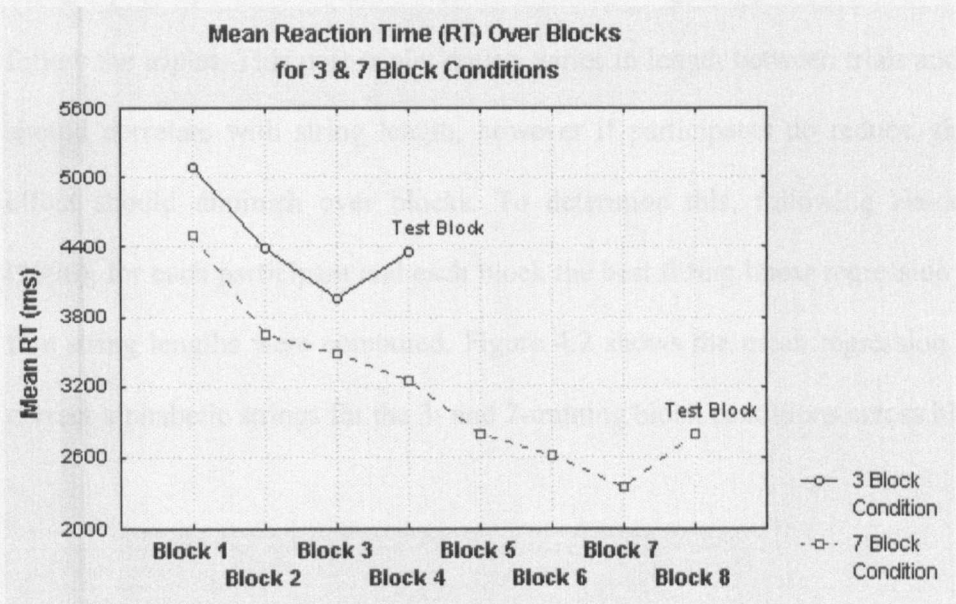
## **4.2.2 RESULTS**

Following the procedure of Haider and Frensch (1996) and the overall methodology set out earlier, the results have three main sections: first the effect of string length on Response Time (RT) for correct and incorrect strings, next the effect of introducing errors into the redundant parts of the strings, both for error rates and RT, and finally a consideration of residual processing within the redundant string items.

Mean error rates were computed for each subject in each practice block, subjects were excluded from the analysis if their error rate was higher than 10% throughout the training blocks (N = 1). This resulted in 12 subjects in the three training block condition and 11 in the seven training block condition. Mean error rates across the two conditions ranged from 1 to 9%

REACTION TIME (RT) DATA

There was no evidence of a speed-accuracy trade-off. The correlation between RT and mean percent error rate for the 7-block group was  $r(N=11) = 0.008, p>0.05$  (excluding responses to strings where the error could occur in the redundant segment) and  $r(N=12) = 0.18, p>0.05$  for the 3-block condition. Figure 4.1 illustrates the mean RT for responses to all strings for 3- and 7-training block conditions.



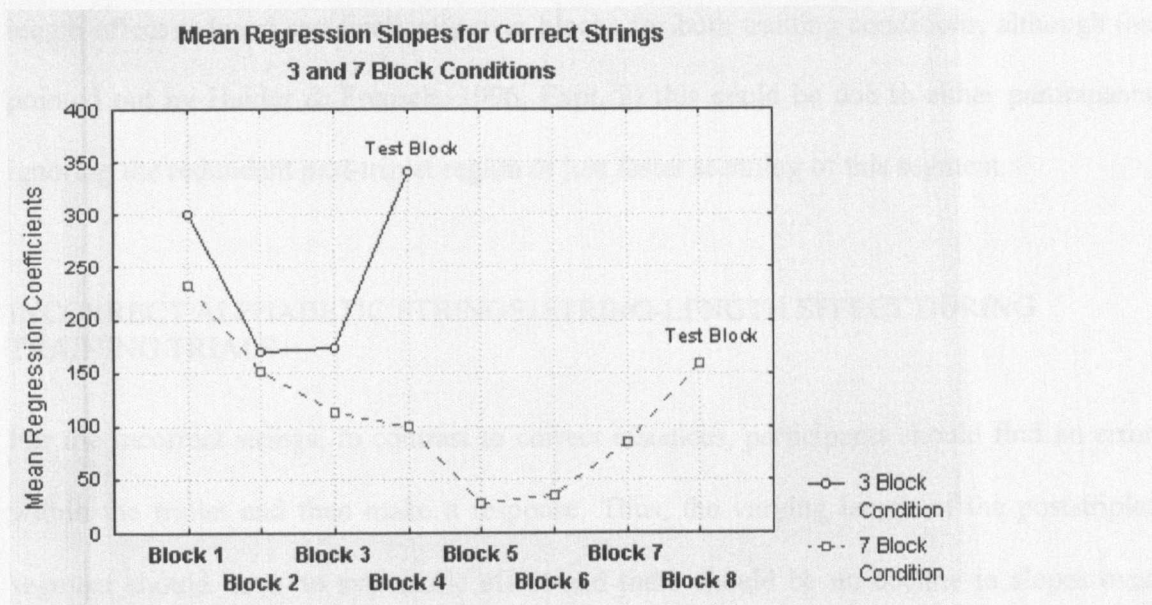
*Figure 4.1: Mean Response Time (RT) for string verification over blocks for both 3 and 7 block conditions for Experiment 1.*



As can be seen, RT reduces over blocks for both 3- and 7-block conditions, A One-Way ANOVA (Block 1-4, within subjects) reveals a main effect of Block  $F(3,33)= 19.19$ ,  $p<0.001$  for the 3-Block condition. Post-hoc tests (Tukey), found a reliable decrease in RT between the first and last training blocks  $p< 0.001$ . Similarly, a block main effect was found for the 7-block condition (Block 1-8, within subjects)  $F(7,70)= 23.29$ ,  $p<0.001$ . Post-hoc Tukey tests again revealed a reliable drop in RT between first and last training blocks. While there was an increase in RT for the test blocks, in neither condition was this increase reliable.

#### CORRECT ALPHABETIC STRINGS: STRING-LENGTH EFFECT DURING TRAINING TRIALS

If IR occurs as suggested by Haider and Frensch (1996, 1999a 1999b), then for correct strings, participants should, over practice, start to ignore the redundant string items that follow the triplet. This post-triplet region varies in length between trials and so initially RT should correlate with string length, however if participants do reduce, this string-length effect should diminish over blocks. To determine this, following Haider and Frensch (1996), for each participant and each block the best fitting linear regression lines across the five string lengths were computed. Figure 4.2 shows the mean regression coefficients for correct alphabetic strings for the 3- and 7-training block conditions across blocks.



*Figure 4.2 Mean regression coefficients over blocks for 3 and 7 training block conditions for correct alphabetic strings, Experiment 1*

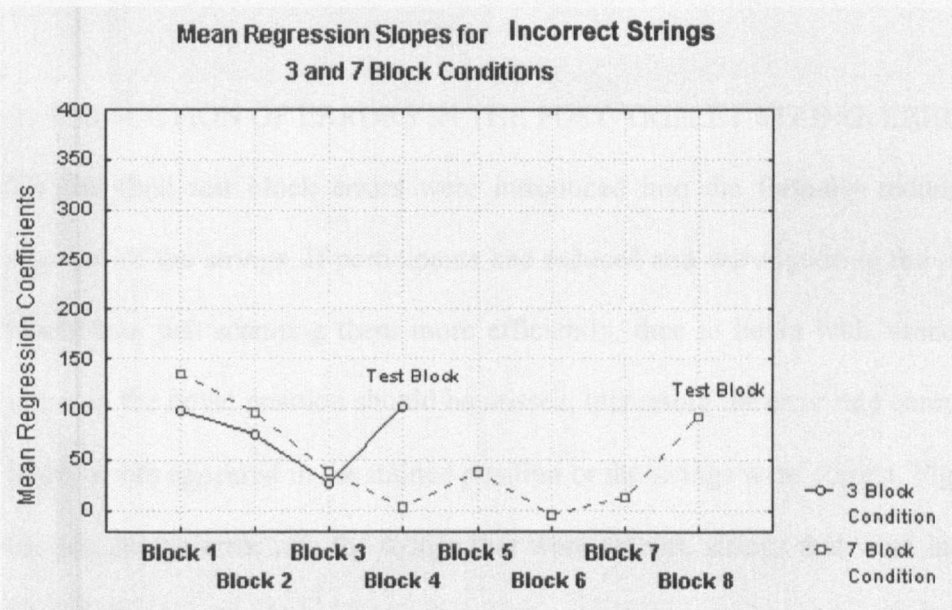
The string-length effect as measured by the regression slopes does appear to decrease over training blocks for both conditions. A 2 (Condition – 3- training blocks & 7-training blocks, between subjects) \* 2 (Training Block – First & Last: within subjects) ANOVA to compare both conditions at these two points, on the first and last training blocks for correct strings indicated a significant effect of Block, First to Last  $F(1,21) = 12.66$ ,  $p < 0.01$ . However, there was no significant effect of Condition (level of training) and no reliable Condition by Block interaction. Thus, there was a similar level of reduction in the string-length effect irrespective of the length of training.

Follow-up One-Way ANOVA with Training Block as a within subjects independent variable for each experimental condition separately (Test block excluded), revealed a main effect of block for both the 3- Block and 7-Block conditions ( $F(2,22) = 4.15$ ,  $p = 0.03$ ;

$F(6,60)= 4.94, p= 0.001$ , respectively). The data are consistent with the view that the string-length effect reduced systematically over blocks for both training conditions, although (as pointed out by Haider & Frensch, 1996, Expt. 2) this could be due to either participants ignoring the redundant past-triplet region or just faster scanning of this segment.

INCORRECT ALPHABETIC STRINGS: STRING-LENGTH EFFECT DURING TRAINING TRIALS

For the incorrect strings, in contrast to correct instances, participants should find an error within the triplet and then make a response. Thus, the varying length of the post-triplet segment should have no systematic effect and there should be no decline in slopes over blocks as found for the correct strings. Figure 4.3 shows the mean regression coefficients for incorrect alphabetic strings for the 3- and 7-training block conditions over blocks.



*Figure 4.3. Mean regression coefficients over blocks for 3 and 7 training block conditions for incorrect alphabetic strings, Experiment 1*

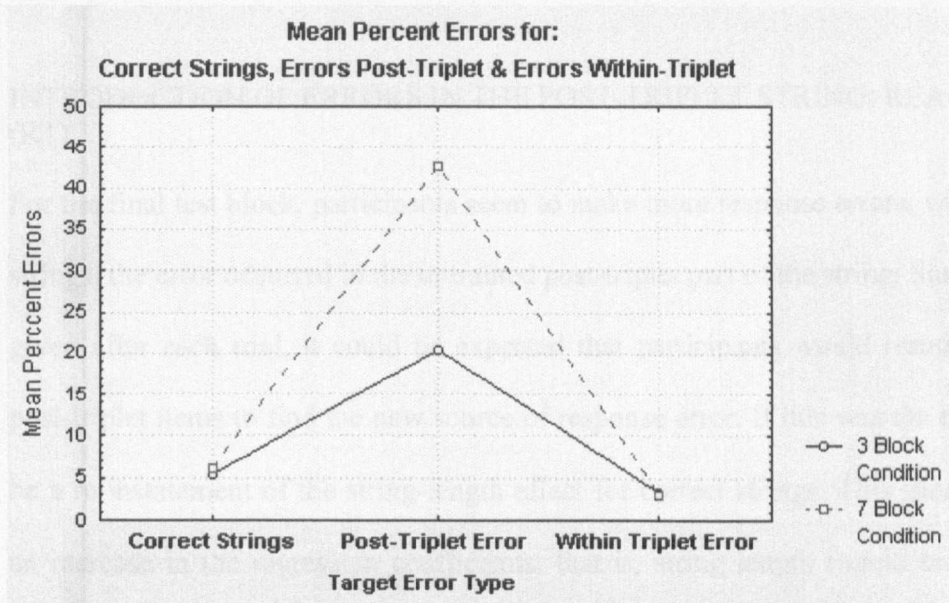
There does appear to be a relatively small decrease in slope over training blocks for both conditions, although the mean coefficients are lower than for correct strings during initial blocks. A 2(Condition – 3-training blocks & 7-training blocks, between subjects) \* 2(Training Block – First & Last: within subjects) ANOVA on the first and last training blocks for incorrect strings found no main effect of Block. Also, there was no significant effect of Condition (level of training) and no Condition by Block interaction.

Follow-up One-Way ANOVAs with Training Block as a within subjects independent variable for each experimental condition separately (Test block excluded), revealed no main effect of block for the 3-Block Condition. For the 7-Block condition, there was a trend for reduced coefficients over blocks, but this did not reach significance. This suggest that the reduction in the string-length effect was much less marked for the incorrect strings, as expected, and starts to approach a zero coefficient after just three blocks.

#### INTRODUCTION OF ERRORS IN THE POST-TRIPLET STRING: ERROR RATES.

For the final test block errors were introduced into the formally redundant post-triplet segment of the strings. If participants had reduced and were ignoring the post-triplet items rather than just scanning them more efficiently, then to begin with, some of these target errors in the novel position should be missed, increasing the error rate compared with when target errors appeared in the trained position or the strings were correct. Figure 4.4 displays the percentage error rate for strings that were correct, strings that were incorrect with the error in the usual position and strings that were incorrect with the error in the novel post-triplet segment.





*Figure 4.4 Mean percent error rate for the final test blocks where error could appear in the trained or a novel post-triplet position; correct string errors added for comparison.*

#### Experiment 1.

As can be seen, there seem to be greater percentage errors in the test block when the target error occurred in the formerly redundant post-triplet position. A 2(Condition – 3-training blocks & 7-training blocks, between subjects) \* 3(Position – No error, Error Post-Triplet & Error Within-Triplet) ANOVA revealed a significant effect of Position,  $F(2,20) = 17.05$ ,  $p < 0.0001$ , but no effect of Condition or Condition \* Position interaction. Post-hoc tests (Tukey), reveal that the higher error rate for post-triplet errors compared to both within-triplet errors and correct strings is reliable for the 7-training block condition ( $p = 0.01$  and  $p = 0.004$  respectively) but not the 3-training block condition. Overall this suggests that particularly for the 7-block condition, more alphabetic errors are missed when they appear in the novel post-triplet position, suggesting that participants are learning to ignore this part of the string and not just processing it more efficiently.

## INTRODUCTION OF ERRORS IN THE POST-TRIPLET STRING: REACTION TIMES (RT).

For the final test block, participants seem to make more response errors, when for incorrect strings, the error occurred in the untrained post-triplet part of the string. Since feedback was given after each trial, it could be expected that participants would resume searching the post-triplet items to find the new source of response error. If this was the case there should be a re-instatement of the string-length effect for correct strings. This should be shown by an increase in the regression coefficients: that is, string length should once again predict RT. The change in slope shown earlier between the last training and test blocks for correct strings in figure 4.2, were subjected to a 2(Condition – 3-training blocks & 7-training blocks, between subjects) \* 2(Block Type – Last Training & Test: within subjects) ANOVA. There was a significant main effect of Condition  $F(1,21)= 12.21$ ,  $p= 0.002$  and Block Type  $F(1,21)=17.74$ ,  $p< 0.001$ , but the Condition \* Block Type interaction did not reach significance.

Follow-up One-Way ANOVAs within each experimental condition, once again using block as a within subjects independent variable (this time including the test block), revealed a significant main effect of block for the 3-training block and 7-training block condition ( $F(3,33)= 4.94$ ,  $p= 0.0061$  and  $F(7,70)= 4.12$ ,  $p= 0.0001$ , respectively). Post-hoc tests (Tukey) found that for the 3 training block condition, the test slope increased significantly from the second and last training blocks ( $p= 0.022$  and  $0.025$  respectively), but did not differ reliably from the first training block. Similar post-hoc analysis found that for the 7-training block condition, the test block did not differ reliably from the last training block. However, further inspection reveals that the last three training blocks do differ significantly

from the first training block ( $p = 0.001, 0.002 \text{ \& } 0.049$ ), indicating the reduction in string-length effect, but the first training and test blocks show no reliable difference, suggesting to some extent the string-length effect was re-emerging.

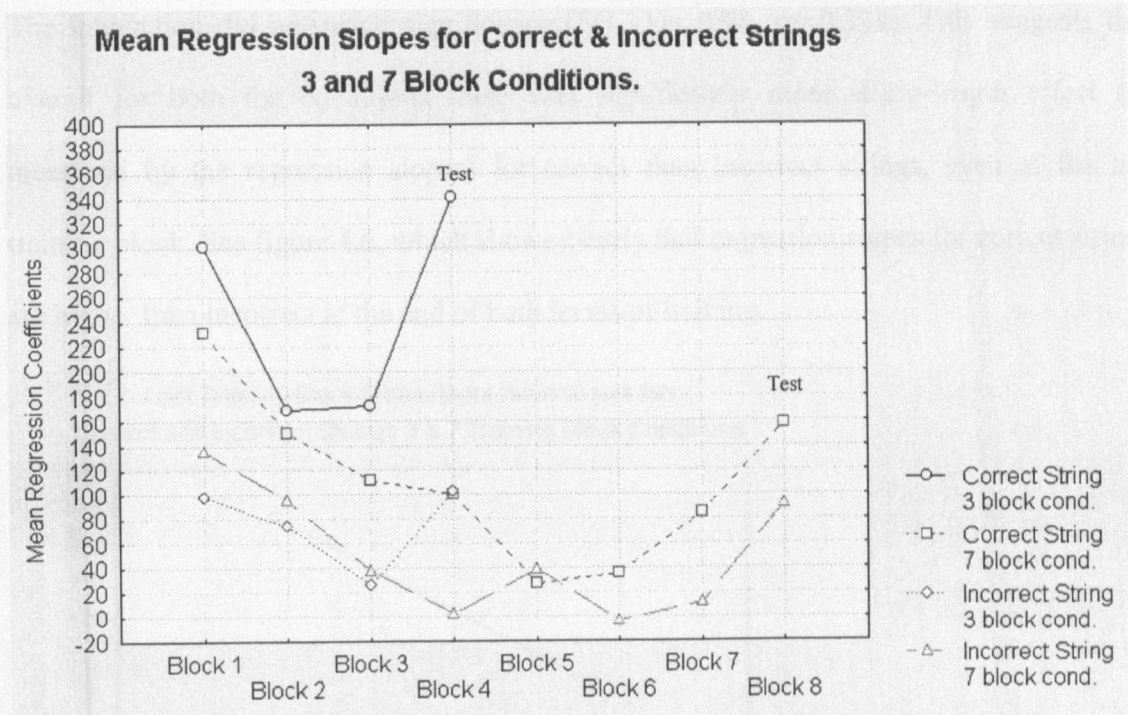
Overall, the results for the response times during the test block suggest that participants were once again starting to search the post-triplet items for errors. The re-emergence of the string-length effect is more convincing for the 3 training block condition than for the 7 block condition, as there is a significant rise in slope between the last training block and the test block for this group. This suggests that participants were, at test, scanning the post-triplet region enough to significantly increase the string-length effect compared to the previous training block. To some extent this concurs with the test block error rates found. That is, participants in the longer training session searched the post-triplet items less reliably during the test block than the shorter training group and so missed more post-triplet errors and subsequently had the highest error rates for the test block.

#### RESIDUAL PROCESSING OF REDUNDANT POST-TRIPLET ITEMS.

It may also be interesting to compare the regression slopes of correct and incorrect strings across blocks simultaneously: the difference between the two may indicate the degree to which IR was occurring. To explain, for correct strings, initially the whole string should be searched for alphabetic errors. However, once the redundancy was realised, only the relevant triplet items should be scanned. In contrast, for the incorrect strings, participants once they found the error following the brackets, could respond immediately never needing to scan the post-triplet area even from the first trials. The data tend to support this, though

there is some evidence of a string-length effect for incorrect strings during initial trials and the test block, possibly due to some confusion of the exact task requirements.

However, after some practice it could be hypothesized, that once IR occurred, correct and incorrect strings should show similar mean slopes for the same block. For incorrect strings it should be realised that no post-triplet scanning was required past the letter-number-letter sequence and for reasons of redundancy, the same should be true following practice for the correct strings.



*Figure 4.5 Mean regression coefficients for correct and incorrect alphabetic strings during training and test blocks for 3- and 7-training block conditions, Experiment 1*

The slopes for the incorrect strings could be regarded as a baseline condition representing what complete IR should look like. Any difference between the slopes for the two string types therefore, may indicate some residual processing of post-triplet correct string items.



Figure 4.5 displays the mean slopes for correct and incorrect strings together for both levels of training condition.

As can be seen, there appears to be larger difference between correct and incorrect strings for the 3 block condition than for the 7-block condition. Correct and incorrect mean slopes for both conditions for the *final* training block were subjected to a 2-Way ANOVA 2(Practice Length – 3 or 7-blocks, between subjects) \* 2(String Type – Correct or Incorrect, within subjects). This revealed only a main effect of String Type  $F(1,21)= 8.4, p=0.009$ . The interaction did not reach significance ( $F(1,21)= 0.98, p= 0.334$ ), This suggests that overall for both the conditions there was significantly more string-length effect (as measured by the regression slopes) for correct than incorrect strings, even at the last training block. See figure 4.6, which shows clearly that regression slopes for correct strings are higher than incorrect at the end of both levels of training.

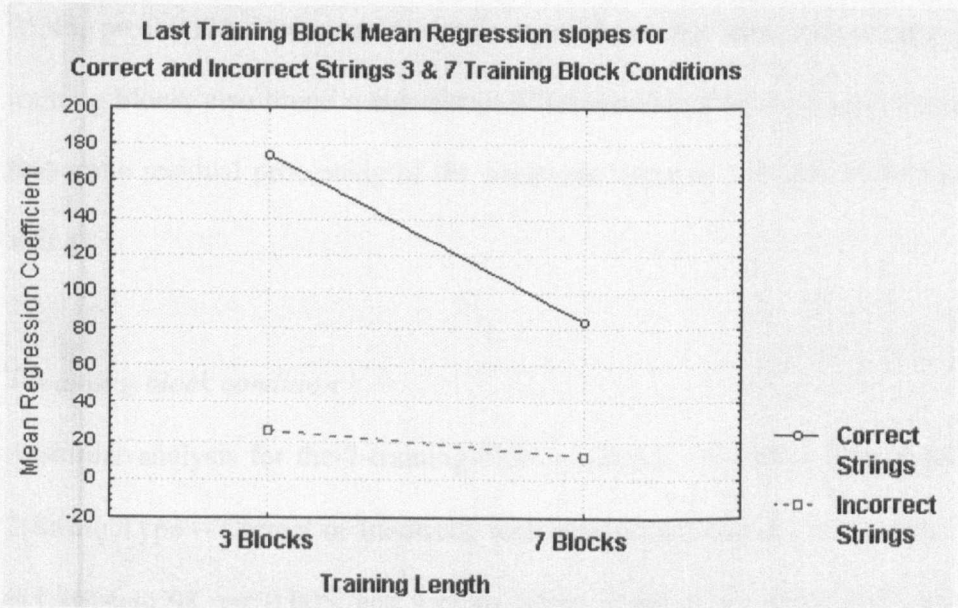


Figure 4.6. Mean regression coefficients for correct and incorrect strings for both levels of training (3 training block and 7 training block), Experiment 1

### *3-training-block condition*

Follow up analyses were conducted for each experimental condition separately, with training Block as the within subjects independent variable. For the 3-block condition a 3(Block – 1 to 3, within subjects) \* 2(String Type – Correct or Incorrect, within subjects) ANOVA found a main effect of String Type  $F(1,11)= 11.77$ ,  $p= 0.006$ , no effect of Block  $F(2,22)= 2.58$ ,  $p= 0.099$ , and no String Type \* Block interaction.  $F(2,22)= 0.89$ ,  $p= 0.43$

The hypothesis was that there would be a difference between regression coefficients for correct and incorrect strings during initial blocks, but this should reduce by the last training block. Particularly, as IR occurred both string types should approach equivalence. To investigate further, planned comparisons were carried out at each of these times. For the 3-training block condition, a comparison of correct and incorrect slopes, collapsing across the first two blocks, found, as expected, a significant difference between string types  $F(1,11)= 25.45$ ,  $p= 0.0004$ . However, comparison of slopes for in/correct string types for the last training block, also found a significant difference  $F(1,11)= 9.066$ ,  $p= 0.012$ . This suggests that some residual processing of the irrelevant segment was still occurring for the correct strings.

### *7-training-block condition*

A similar analysis for the 7-training block condition, 7(Block – 1 to 7, within subjects) \* 2(String Type – Correct or Incorrect, within subjects), found a main effect of String Type  $F(1,10)= 11.98$ ,  $p= 0.006$ , and a main effect of Block  $F(6,60)= 5.92$ ,  $p< 0.0001$ , but no String Type \* Block interaction  $F(6,60)= 0.65$ ,  $p= 0.69$

Planned comparisons between string types once again with the first two initial blocks collapsed finds a significant difference during early training,  $F(1,10)= 13.40$ ,  $p= 0.0043$ , however, comparison of String Types for the last training block this time finds no reliable difference  $p>0.05$ . This suggests that there was no reliable difference, in the string-length effect as measured by regression slopes, between string types after seven training blocks.

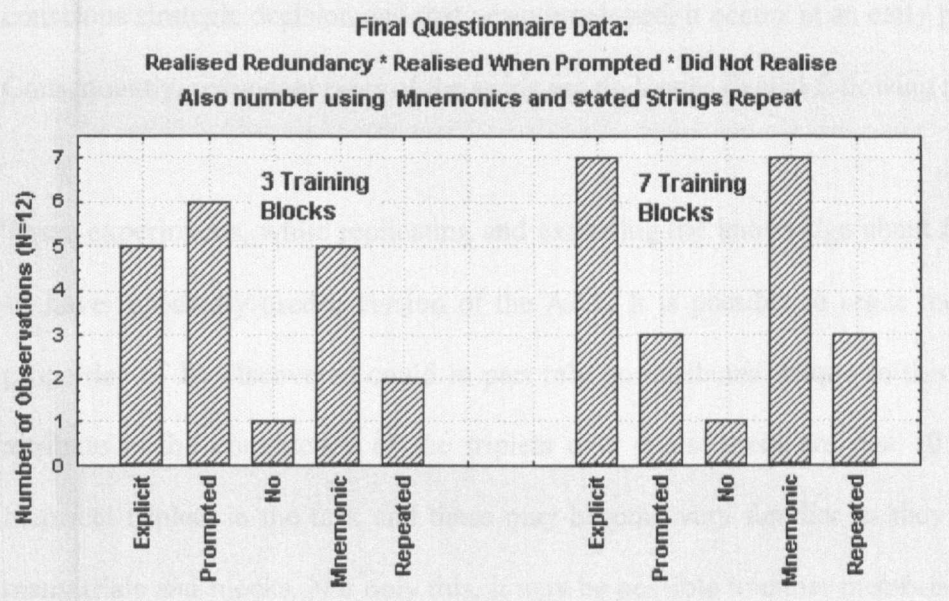
The difference in the regression slopes between string types may reflect the level irrelevant items were processed at during the task. Accepting this, it seemed that increased training brought a following reduction in this residual processing. After seven training blocks, there appeared to be very little residual processing of the post-triplet segment.

#### END OF SESSION QUESTIONNAIRES

Following the completion of the string verification trials, each participant was asked a number of questions (see appendix 2). The questions began as just a vague enquiry if anything unusual was noticed about the strings to more specific questions about the position of the alphabetic errors. The aims were first, to determine whether the underlying ‘rule’ and redundancy was consciously noticed and second, to find out whether any particular strategies other than IR had been employed. Figure 4.7 shows for each condition the number of participants that explicitly stated knowledge of the redundancy without prompting, those that realised it when they were prompted by the suggestion of the redundancy and those that had not realised the rule at any level.

For the 3-training block condition only one person said they had no idea of the rule even after they were prompted that this may be the case. This was also found for the 7-training

block condition. Of course the prompted responses could to some degree reflect demand characteristics with participants making an affirmative response to the question. More interesting was the number of participants that reported the rule, indicating the post-triplet segment was redundant, without prompting. As can be seen after seven blocks more participants explicitly knew the rule compared to after three blocks. The reverse was true for the prompted responses, which could suggest, bearing the earlier caveat in mind, that participants were beginning to know the rule at some level after three blocks, but were only confident in stating it after further practice.



*Figure 4.7 Participant responses to final questionnaire: Those that explicitly realised post-triplet redundancy, Those that realised it when prompted and that did not know of the rule. Also shown is the number that reported using mnemonics and thought the strings repeated, Experiment 1*

Across both conditions twelve of the twenty three participants reported using mnemonics and five said the strings repeated. Overall this suggested that the case-alternation manipulation was not successful in making the strings less familiar or memorable.

### **4.2.3 DISCUSSION**

Haider and Frensch (1996) found that participants could learn whether a task contained any redundancy and then ignore the redundant elements, so aiding task performance. Later studies (Haider & Frensch, 1999a; 1999b), extended these findings, suggesting that IR is a conscious strategic decision and that when employed, it occurs at an early perceptual stage. Consequently, redundant parts of the string are no longer fixated following practice.

These experiments, while replicating and extending the knowledge about the principles of IR have repeatedly used a version of the AAT. It is possible to argue therefore, that the properties of IR discovered could in part rely on attributes unique to this task. One such attribute is the consistency of the triplets over trials: there are just 10 correct and 10 incorrect triplets in the task and these may become very familiar as they repeat over the many trials and blocks. Not only this, it may be possible to either memorise the triplets, or form mnemonics out of the letter pairs either side of the digit.

One aim in running the present study was to test whether adding random ‘noise’ in the form of mixed letter-case would affect IR over practice. Thus making the strings less familiar over trials and mnemonics less likely to be adopted. If IR is involved in skill acquisition (Haider & Frensch, 1996; 1999a; 1999b), it could have implications for applied psychology

As such, IR should be robust enough to occur in naturalistic situations as well as more controlled experimental settings, so the effect should persist even if the stimuli are less consistent or ‘noisy’. IR under these conditions further suggests it is an important component for inclusion within contemporary theories of skilled performance.

### *Evidence of information reduction*

The results from the current study support the earlier central findings of Haider and Frensch (1996, Expt.2), suggesting that IR can occur within a task without explicit instruction to the participant to ignore the redundancy. The experiment reported here mostly replicates this earlier experiment: finding a reduction in mean regression slope over blocks: indicating the post-triplet segment has less effect on response time with increasing practice. Also the two experiments concur in finding an increase in response errors when the alphabetic error occurred in a novel position during the final test block. The convergence of a reduced string-length effect (as measured by the regression coefficients) and the increase in post-triplet errors during test, strongly suggest that participants were ignoring the redundancy rather than just processing it more quickly.

As found earlier, there is also evidence that IR increases with practice: the mean regression coefficients are of a much lower value for the final training block in the 7 block condition than for the 3 block condition. Further support for this comes from the test block errors. A higher percentage of post-triplet alphabetic errors were missed in the 7 as opposed to 3 block conditions. It appears then, that IR increases with more blocks, and there is some indication that it has occurred by around 500 trials, which is in agreement with Haider and Frensch’s (1999a, pp131.) earlier view describing the progression of IR.

Overall, the present experiment therefore finds IR to occur. It seems also that adding variation to the strings in the form of random letter case-mixing over trials did not disrupt the development of IR. There are two possible conclusions at this point: either stimulus familiarity has no influence on the selection of an IR strategy, contrary to Schunn (1997), who found strategy selection was effected by such a dimension. Alternatively, while changing letter-case may have made the strings more ‘noisy’, the manipulation did not really disrupt stimulus familiarity as hoped.

#### *Effect of case-mixing*

Taking this last alternative first, there is evidence from other areas that case-mixing disrupts recognition performance (Cole & Haber, 1980; Mayall *et al*, 1997; Posner, 1967; Posner & Keele, 1967; Posner *et al*, 1969). Admittedly this concerns mainly letter and word recognition in lexical tasks. However, the strings certainly were perceptually different as the case alternated randomly from trial to trial.

Additionally, There is some evidence from the implicit learning literature that unusual characters may affect the mere exposure effect (MEE), if not implicit learning itself (Zizak & Reber, 2004). MEE it seems results from feelings of familiarity (See Borstein, 1989, for an early review) and so unusual character formats may moderate this sense of familiarity. One could reasonably propose that strings of case-alternating letters can be regarded as an

unusual format. So there did seem grounds to believe that mixing letter-case would reduce string familiarity over practice.

However, the self-report questionnaires seem to suggest participants were still engaging in the formation of mnemonics and did realise the strings repeated. This is probably because the strings were processed beyond a purely perceptual level and the letter names were necessarily accessed to make the triplet calculation. Subsequently, at the letter-name level the strings did become familiar. It seems then, that the familiarity manipulation was not successful, although IR has been shown to persist even when there is some stimulus variability over trials. This certainly increases confidence in the robustness of the effect and the likelihood it can occur in more naturalistic situations.

### *String Familiarity*

While Haider and Frensch (1996 Expt. 3) have shown IR is not stimulus-specific via transfer tests, at the risk of labouring the point, string familiarity is still interesting for a number of reasons. While IR may not be stimulus-specific once adopted, realising the strings are repeating may be necessary for it to develop. It may be that some form of over trial consistency is necessary for participant to gain the confidence to ignore part of the stimulus.

### *Mnemonics*

It does seem that there is still much to learn about the mechanism underlying the detection of environmental regularities such as the consistency built into the strings here. Firstly, it is an empirical question whether IR ‘develops’ at an explicit or implicit level, though



Frensch *et al* (2002) do seem to be regarding it as an implicit process. Secondly, participants in this study do seem to learn the triplets, even making mnemonics out of the letter-pairs either side of the digit. The reason for this was obviously to avoid the fairly difficult and repetitive triplet calculation. Nonetheless, it must quickly become apparent that the same triplets were appearing quite regularly, even if the post-triplet segment varied in length. As feedback was immediate, and responses to triplets that were repeating always gave positive results, this may play a role in the confidence to adopt the reduction strategy.

The retrieval of triplets or letter pairs from memory is interesting, as the whole point of Haider and Frensch's (1996 Expt. 3) transfer experiment was to set-up IR against Logan's (1988) instance theory. This theory posits that task performance improves (to an automatic level) as many instances of a task become stored, to the point that retrieval of a correct instance 'wins' out in processing speed over any algorithmic or calculation method. A method that should become unsuccessful following transfer to a new stimulus-set.. It would seem though that in the AAT participants were searching for strategies to allow just such an instance retrieval process to occur and so avoid the triplet calculation. While this is not exactly what Logan had in mind, it does suggest that 'instances' are playing a role in the task. Unfortunately, the case-mixing manipulation used in this current experiment did not disrupt this mnemonic strategy, so the role of instance retrieval in the formation of IR cannot be determined at this point. A further experiment increasing the number of unique instances, would prove interesting to further this investigation.

### *Length of practice*

Whilst the case-mixing did not seem to prevent the learning of triplet instances and the current results were in line with earlier published accounts of IR, there were though, some differences between the results of the earlier (Haider & Frensch, 1996, Expt. 2) and current experiment. There did seem to be a significant reduction in the string-length effect after three training blocks in the current experiment and a subsequent significant increase in the same for the test block. The earlier published study did not find this until five training blocks had occurred. However, for the 3-block group, the rise in incorrect responses for post-triplet target errors during the test block was not reliable, which may indicate reduction was not that complete after all. In comparison, the 7-block condition in the present experiment did result in both a significant reduction in the string-length effect over training blocks and a reliable increase in incorrect responses for post-triplet errors during test. However, this time the test block increase in the string-length effect was not reliable, suggesting perhaps that there was still some reduction taking place.

In contrast to this, Haider and Frensch (1996, Expt. 2) report that the increase in string-length effect during test was a function of practice, with the greatest increase from last training block to test occurring for the 7-block condition. They reason that when it was realised that errors were occurring in the post-triplet segment, flagged by feedback indicating incorrect reposes are being made, the formally redundant segment will start to be scanned again, and the string-length effect re-instated. They further suggest that, at a group level, this increase between training and test should be greater for conditions with much practice compared to conditions with little practice (Haider & Frensch, 1996, pp 325).

While this is what they found, there was no reason to ‘expect’ this should be the case. While it is true there is more potential for increase after longer training: the coefficients (which indirectly measure the string-length effect) drop lower with practice, so there is more room for reversal. There is also the possibility that reducing more completely and for longer will be more difficult to ‘break out of’ in longer compared to shorter training conditions. If the results of the current experiment are viewed from this perspective they do seem to make some sense.

In the present experiment, for the 3-block condition the slopes dropped less then increased significantly during test. Suggesting participants initially reduced slightly, but quickly stopped this at test and so made very few post-triplet errors. In the 7-block condition the slope for correct strings decreased more, but did not rise reliably at test, suggesting participants were still trying to reduce to the triplet. Subsequently, many more post-triplet errors were made in the test block. It seems logical that the greater the reduction that remains during the test block the more post-triplet errors that will be made, which is exactly what was found in the present study. There does seem to be evidence supporting both views: whether the point at which the post-triplet segment starts to be processed again during the test block is moderated by practice or not, is still an empirical question requiring further experimentation.

### *Residual processing*

Haider and Frensch (1996, 1999a, 1999b), do not analyse the regression slopes for correct and incorrect strings together, possibly because they have also found a decrease, although attenuated, in the string-length effect for the incorrect strings. Logically this should not

occur because once the triplet error is found in an incorrect string, no further scanning of the post-triplet string is needed, since there was always only one error. This should mean that for the incorrect strings, varying the number of post-triplet letters should not effect response times systematically. The current experiment though, also found a marginal decrease in slopes over blocks for this string type. It can only be assumed that this was the result of some confusion during initial blocks and possibly over cautious checking.

Nonetheless, once settled into the task, the slopes for the incorrect strings should represent a baseline, indicating what complete reduction should look like. The difference between correct and incorrect slopes is interesting then, since it represents how much residual processing of the post-triplet string takes place for correct strings as practice progresses. This should inform if IR is fully adopted, or just partly, with some intermittent checking of the whole string. While this analysis is not entirely convincing, for the reason mentioned above, it does seem to suggest there was significant checking of the whole string following short practice session, but relatively little for the longer practice condition. By seven training blocks there was no reliable difference between the coefficients for correct and incorrect strings. In fact, inspection of figure 4.5 suggests IR had largely occurred by five blocks, in line with Haider and Frensch's (1999, pp131) earlier report. Overall, this implies that after enough practice, IR was almost completely adopted, and the post-triplet string was only scanned more fully when prompted by error feedback during the test phase.

## SUMMARY

The results reported here have replicated the earlier work of Haider and Frensch (1996, Expt.2) in finding what they term Information Reduction (IR). Once again participants

learned over practice to isolate and ignore the in-built task redundancy. So reducing task load, but leading to increased errors when the redundant segment later became task-relevant. That this occurred even with the case-mixing manipulation suggests that IR is robust enough to occur in situations less perfect than the typical experimental setting. Further, the effect is also robust enough to find with small sample sizes, once again indicating the effect to be general and frequently occurring.

Some interesting questions are raised over the effects of practice and the return from IR when the task redundancy becomes relevant. However, perhaps more interesting is that a mnemonic forming strategy was found, which although indicating the case-mixing manipulation failed, does pose some questions about the AAT and attributes peculiar to this task. This is particularly interesting since the suggestion is that IR is a form of learning not accounted for by contemporary theories of skill acquisition, particularly instance based accounts (Logan, 1988). Yet we found participants searching to achieve just such instance learning and retrieval.

It does seem possible that some form of string recognition, familiarity or a feeling that strings are repeating may contribute to the development of IR. The most obvious way to test this is to increase the number of unique instances, so that such learning is not available during training phases. Unfortunately, this is not easily achieved with the AAT, so for subsequent experiments a new paradigm will need to be developed.

## **CHAPTER 5: STUDY 2**

### **Information Reduction in a Target Search Task (TST)**

*Previous studies have repeatedly used the alphabet arithmetic task (AAT) to investigate information reduction (IR). This chapter describes two experiments exploring reduction within a completely new task*

## **5.1 INTRODUCTION**

Haider and Frensch (1996; 1999a; 1999b) have argued that information reduction (IR) was applied under conscious control and also that it occurred at an early perceptual level. They have also found that IR was not stimulus-dependent; once learned it persisted across changes in stimulus-set. These findings begin to suggest factors underlying the deployment of IR once it had been learned. However the conditions under which it develops are also interesting.

Some of these conditions have already been set out. IR developed even if explicit instructions informing of the irrelevancies were not given (Haider & Frensch, 1996). Also, the positioning of the irrelevant information was non-critical for learning the strategy (Haider & Frensch, 1999b). So, it seems irrelevant information can be isolated and ignored without any specific instructional cueing that revealed the redundancy. Also, positioning the triplet at the end of the string or alternating its position across trials still allowed IR to occur. This indicates two conditions under which reduction strategies develop. Importantly, IR was learned incidentally, and the position of the relevant information was non-critical. This certainly suggests that IR may develop in many tasks and may be a normal consequence of practice and skill acquisition.

Green and Wright (2003) extend this work further, reporting that not just irrelevant, but also redundant information may be ignored. In their task, when information was duplicated in a string, this relegated one information source as redundant. Specifically, if the triplet was correct then so was the post-triplet and likewise, if one was incorrect then so was the

other. Under these conditions it was the information processed first in the task to which participants reduced. This suggests that when duplicated, later processed information was most likely to become ignored. In summary, it seems IR may develop for both 'irrelevant' and duplicated 'redundant' information. Additionally, reduction occurs without instruction or the consistent positioning of the relevant information. Such factors are important in understanding the conditions under which IR may develop.

However, there is still a great deal to discover about the conditions under which IR is learned. To re-iterate, Haider and Frensch (1999a), propose that IR occurs as a conscious strategic decision; one that can be affected by task instructions. Also, they suggest that IR is not stimulus specific and persists even if the stimulus-set is switched. This indicates the underlying rule was abstracted, rather than the occurrence of some sort of instance retrieval. One argument though, is that these factors play a role once IR has been adopted, but may not apply during initial learning stages.

While once developed, IR may be applied consciously and early on in processing of the AAT strings, the mechanism by which the rule was learned has yet to be specified. Initial learning may thus proceed at an implicit or explicit level. Similarly, while IR may, once formed, persist over stimulus-set changes, one could propose that some level of stimulus consistency was necessary to *learn* the in-built regularity within the strings. It is likely therefore, that a number of factors regarding the development of IR have yet to be identified.



Knowledge of the conditions under which IR may develop are further limited by the repeated use of the AAT in all the studies mentioned above (Haider & Frensch, 1996; 1999a; 1999b). While it is true that IR has been shown to replicate under a number of manipulations of this task, it may also be possible that some attributes of the AAT promote reduction. There is though, some suggestion that IR occurs in other tasks. Lee & Anderson (2001) found what they report as IR in an Air Traffic Control task (ATC). In the task participants are to land planes while adjusting for plane size, runway length, wind direction and fuel load. During the task eye fixation was tracked and linked to screen positions.

Interestingly, as participants got faster at landing the planes, they were less likely to run out of fuel and so monitoring fuel level became less critical. As a result, eye fixations on fuel information became less frequent. Post-hoc Lee and Anderson (2001) attribute this change to IR of increasingly task-irrelevant information. While it is interesting that IR seems to develop for information that starts out as important, but becomes less so, there are some problems with their interpretation. Firstly, the fuel load was neither 'irrelevant' nor 'redundant' within the task: The fuel level was always relevant, even if it became less critical. Neither was it redundant in the sense Green and Wright (2003) proposed: the fuel information was not duplicated elsewhere.

A further problem was that information relevance was not directly manipulated in the Lee & Anderson (2001) study; rather IR was used as a later interpretation for part of the task behaviour. Subsequently, areas of the screen categorised by them as irrelevant were found to be fixated as practice progressed. It is difficult to imagine why whilst engaging in IR, participants would ignore an area that has at least some task relevance, to scan areas that

are assumed to be task irrelevant. It would seem that there is still some clarification needed concerning the development of IR strategies. The mechanism by which people learn which information is redundant and when to ignore it is a complex issue that has only begun to be addressed.

## **5.2 EXPERIMENT 2: THE TARGET SEARCH TASK (TST)**

Whilst there is some indication that IR may generalise to other tasks, further evidence is required. Of particular concern in chapter 4 study 1 was the possibility that mnemonics could develop. This seemed to occur even though the percept of the strings was varied by case-mixing. It would seem that such a strategy is endemic to the AAT and may be necessary for the ‘development’ of IR. Haider and Frensch (1996; 1999a; 1999b) propose IR is not explained by accounts of skill acquisition relying on instance (Logan 1988) or procedural retrieval (Anderson, 1983; 1987). The learning of string instances within the AAT is problematical for this interpretation. Thus the next goal could be to use a novel task that does not allow mnemonics to occur as skill develops.

The original AAT also has another particular attribute, in that the task is divisible into two sub-tasks, the triplet calculation and the post-triplet alphabet recognition (post-triplet items just need to be scanned to ascertain if they follow the alphabet). The former is novel and quite difficult, while the latter is an over-learned skill. Thus, forming two separate and different sub-tasks, one which is relevant and one that is irrelevant to a correct response. To some extent the problem of greater triplet than post-triplet saliency has been addressed in an unpublished study by Lincourt *et.al.* (cited in Haider and Frensch, 1999b) In the study, a

version of the AAT was used in which the strings were made up of a varying number of triplets, such as E[4]J D[4]I G[4]L. Using this method the authors still found IR occurs, indicating that triplet saliency was not necessary for IR.

However, the task was still divisible, segmenting into triplets, and so a number of sub-tasks, both at the perceptual level and during the calculations. It does seem possible that learning the relevant and irrelevant segments may be easier when they form part of different sub-tasks. Subsequently, it is possible that IR would be less likely for a task where no clear demarcation between relevant and irrelevant information exists.

As suggested in the previous chapter, finding an alternative task to the AAT is non-trivial for a number of reasons. Mainly, the task has to be novel so all participants start from a similar level of competence; also the task has to be mastered in a short experimental session. Finally the task has to incorporate a built-in regularity that facilitates IR and leads to a performance improvement.

There is though some indication that IR may arise in different task domains: Haider and Frensch (1996; 1999a; 1999b) have summarised the traditional belief in many areas of psychology that information needs to be selected for action. Such a view is found in literature on perception (Gibson & Gibson 1955, Neisser & Becklen, 1975), concept formation (Regeher & Brooks, 1993), educational psychology (Bransford, Sherwood, Vye & Rieser, 1986) and expertise (Shapiro & Raymond, 1989), so a search through the tasks used should provide a suitable candidate to further test the IR hypothesis.

### **5.2.1 TARGET SEARCH TASKS**

#### **REDUCED PROCESSING AND AUTOMATICITY**

In the area of automaticity, there is a hint that strategies occur to reduce the information processed. For example, in the well known Target Search Task (TST) used by Schneider and Shiffrin (1977) a memory-set of letters (1,2,3 or 4 consonants) was learned, after which a string of 1-4 items (consonants or numbers) was presented and the participant had to decide as quickly as possible if one of the memory-set was present or not.

For example if the memory-set was H B K D

And the visual display was 4 3 B 7,

the answer would be 'yes' a memory item was present.

The crucial manipulation was the stimulus mapping between memory-set and display items. In their constant mapping condition only consonants were used as memory-set items and only numbers were used as distracters in the display set (or vice versa). By contrast, in the varied mapping condition, both memory set and display distracters were a mixture of numbers and consonants. They found a large performance difference in response times between the two mapping conditions. Also, increasing the number of items in memory and display sets had little effect in the constant mapping condition, whereas set size strongly affected reaction time in the varied mapping condition.

While Shiffrin and Schneider (1977) interpreted this as the difference between automatic and controlled processing, Cheng (1985) extends a different explanation. Attributing the

performance in the constant mapping condition not to automatic processing, but rather to the selection of a particular method. Participants realising that any consonant in the display set would be a target, just looked for any consonant amongst the numbers and did not have to search the memory set for a match. Thus, the task became one of categorization; for the constant mapping condition the memory set was given a 'category tag' and this was used to scan the display set. In this way, the amount of information processed (the memory set) was reduced. It seems possible then, that some adaptation of a TST may provide a novel IR paradigm.

## REDUCED PROCESSING AND CUES

Staying with visual search, there are a number of experiments that suggest past experience or cues can encourage heightened selectivity and improved perceptual performance. Perhaps the most well known evidence of this comes from a study of expert radiographers by Carmody, Nodine and Kundel (1980). Skilled x-ray interpreters work very quickly but with great accuracy. This is even more surprising when observation reveals they often seem to ignore abnormalities within the image. It would seem that past knowledge is brought to bear in decisions where to look and what to look for in an image, whilst simultaneously ignoring spurious information.

The cueing of relevant information also seems to improve threshold perception for movement. Ball and Sekuler (1981) determined that the ability to detect very faint moving dots on a video screen could be heightened by a preceding cue; a line that indicated the direction of dot movement. If this cue misdirected attention though, performance was worse than with no cue at all. Certainly it seems that the cue, when present and appropriate,

lowered the threshold at which the dots were perceived. Both these above experiments suggest that any cue that points to relevant information aids performance, whether that cue comes from past experience or some explicit prompt in the stimulus display. This improvement seems to come mainly from some enhancement in the selectivity during the allocation of attention.

Another form of signal that aids visual search is contextual cueing. This phenomenon is perhaps even more interesting as far as IR is concerned, as it appears that context can be used to reduce the search time for a target amongst a display of distractors. That is, part of the display may become ignored under certain contextual cues. To investigate this, Chun and Jiang (1998, 1999) and Olson and Chun (2001) ran a series of experiments where the target was consistently paired with a certain context or spatial pattern of distractors that repeated over trials.

The typical task used these spatial context cues; the pattern of distractors in a display (Chun & Jiang, 1998). Each of 30 blocks contained 12 old and 12 new displays randomly presented. The target was a rotated *T* amongst a distractor display of rotated *L*'s. The twelve old displays were twelve particular configurations of these distractors each linked to a certain target location. The important point was that these twelve old displays repeated across the 30 blocks, whereas the new displays were random across blocks. So in effect the old displays were predictive of target location while the new displays were not. A significant enhancement in search speed was found for the old compared to new displays.

To account for this Chun and Jiang (1998) suggest that each of the repeated instances must have been learned and then discriminated from the novel configurations. This information was then used to guide attention to the target location, so speeding search time for the old displays (in Jimenez, 2002, p.279). Again this provides some evidence that the information processed may reduce within a target search,; prior experience may be used to guide attention to a particular target location without scanning the whole display. So, such tasks may be open to IR *per se*.

In the same study (Chun & Jiang, 1998), only 20% of participants realised the repetition. When given a recognition task, discrimination by participants between old and new displays was at chance. This suggests that their knowledge was at an implicit level. This was further supported by later tests, that indicated participants were unable to guess the target location (after it had been replaced by a distractor) when presented with an old display. Also, it appeared that informing participants of the repetitions with instructions to attend to them hinders cueing, which further suggests that the underlying knowledge of the effect was at an implicit level. Overall, it does seem that contextual knowledge, albeit implicit, can be used to constrain attention or alternatively to prime a response to a relevant region of the display to speed task performance. Whichever, this does seem very close to a form of IR, and again suggests that some form of Target Search Task (TST) may be a useful paradigm to investigate IR.

## APPLIED TARGET SEARCH

From an applied perspective, target search and vigilance tasks have ecological validity, often occurring as part of everyday events, from scanning text for items of interest (Rayner,

1998) to driving down the road. Underwood, Crundall and Chapman (2002), suggests visual search is an aspect of driving that is crucial to understanding and describing the abilities of the skilled driver. Such search strategies are vital both for navigating a route and for the avoidance of hazards and other vehicles. Underwood *et. al.* (2002) also report that one of the single most common reasons for automobile accidents was a failure to search the roadway, more specifically a failure to search far enough ahead or general inattention or distraction. Often, it was younger drivers who suffered most from these problems. Certainly in this country there have been many T.V. programs encouraging drivers to be extra vigilant. Particularly it seems when looking out for motorcyclists; car drivers often report that they did not see them when driving out at a junction.

On a more work-related front, quality control often includes scanning for errors in a product and this may become more critical when safety is an issue, as in aircraft maintenance. Understanding target search has obvious relevance to air force and military applications, Diaz, Hancock and Sims (2004) stress this point. They explain that visual search is an important part of any tactical combat mission. The search for targets becomes particularly relevant when the objects to be sighted are moving or hiding; all of this of course has to be done quickly under combat conditions. They have found that search effectiveness can be moderated by stress and object shape (Diaz *et.al* 2002. Gibson, *et. al.* 2000). It is possible that under these stressful conditions, IR may also moderate the way such searches are undertaken.

Attentional narrowing certainly seems possible under battlefield stress, Hancock, Szalma and Weaver (2004) report such a possibility. Building on the earlier work by Easterbrook



(1959), they demonstrated that the range of environmental cues utilised reduces under emotional arousal: Initially irrelevant cues are eliminated, but eventually, with increasing stress relevant cues also start to become sacrificed in order to maintain performance. Such high-pressure conditions are likely in a military setting. Lieberman *et. al.* (2004) again describes deficits in cognitive function, one of which is vigilance. As Brill *et. al.* (2004) points out; the army is using increasingly advanced technologies which may provide amounts of information greater than the human cognitive capacity to deal with it. Subsequently, understanding the mechanisms underlying visual search for relevant information becomes ever more important.

Aircraft systems are also increasing in complexity and automation. With multi channel instruments and alarms (Mouloua *et. al.*, 2004), scanning such arrays becomes progressively more difficult, especially when a dangerous fault occurs. Under fatigue stress Russo *et al.* (2004) have found that pilots neglect peripheral information under certain simulated conditions, again indicating that attention narrows under duress. Thus it would seem for the modern military, understanding how visual search takes place and how it breaks down will become increasingly important as an area for research. This is even more so when the number of areas a visual search is conducted in everyday life is considered.

## TARGET SEARCH AND IR

Understanding 'visual search' maybe important, but it is an empirical question as to whether IR moderates such search strategies. Nonetheless, it does appear that a TST may prove a fruitful and valid avenue in which to explore IR. It was encouraging that searching for a target in both the Schneider and Shiffrin (1977) and Chun and Jiang (1998) tasks

seemed to lead to some reduction of the information processed following practice. In the former, it seemed likely that the memory set was no longer scanned after some experience with the task, whereas in the latter, areas of distracters become ignored. This last point is particularly interesting as it implies that during a target search, processing may be constrained to a relevant area of the stimulus display. A form of target search where the target appears in an invariant position could therefore allow IR to develop without the need for contextual cueing; the cue would become the regularity of the target position.

However, repeating a display of distractors in various configurations with the target always at the same position would not be much of a test, and would hardly constitute IR, rather it would demonstrate instance learning (Chun and Jiang. 1998). A version of the Schneider and Shiffrin (1977) TST however, makes for a more viable task. Similar to their paradigm, a target memory-set could be given that was searched for in a display-set, both increasing difficulty and adding apparent task validity. For example: -

If the memory-set was D P W B

And a display-set was A W V X

This requires an affirmative answer, as the 'W' from the memory-set is present

For a following display-set of S B R Y

A negative response is required as none of the memory-set are present.

The in-built regularity to allow IR in this case could be that, if present, a target always appeared in the same position in a string. Whilst seemingly trivial, such a task is in many

ways analogous to the AAT; initial instructions are to search the whole stimulus, but only part of that stimulus is relevant to task completion.

## OVERVIEW OF THE TARGET SEARCH TASK (TST), IR PARADIGM

One goal in generating a new task was to demonstrate some generalisation of IR beyond the often used AAT paradigm. To be successful in this, the task needed to be both cognitively and perceptually different to the original task. In designing the new task, two clear aims were in mind. First, it had to be seen as a holistic task. That is, it should not break down into sub-task that were relevant and redundant to correct performance. This should inform whether some pre-existing segmentation of these information types was necessary for IR to develop in the AAT. Second, the number of unique task instances should be much greater than the AAT, so that learning of particular repeating instances could not take place. This should then rule out any instance account describing IR during its development.

This last point is particularly important. Haider and Frensch (1996) posit that IR is not explained by instance accounts such as Logan's (1988). Not only this, but learning of the instances appeared to be endemic in the AAT. Some participants previously (Study 1, Chapter 4) reported the use of mnemonics to aid recall of particular instances. Considering this, it does seem likely that some instance learning may be necessary for IR to develop, even if it is not stimulus-specific at a later stage. Logan's (1988) instance theory has been a milestone in the understanding of skilled performance, so this is an key property to examine further. Some form of TST should achieve both this and the previous aim set out above. Any number of display strings may be generated making instance learning unlikely. Also,

there needs to be just one instruction to search the whole string, with no sub-tasks within this.

As suggested earlier, the task was designed to take the form of a search-task similar to that used by Schneider and Shiffrin (1977) to investigate automaticity. Unlike their task though, both the memory-set and the display-set consisted of letters. The to-be-remembered memory-set included three letters chosen randomly from approximately the beginning, middle and end of the alphabet, the display-set was a randomly generated string of letters from 3 to 7 items in length. These string lengths were selected to be comparable with those in the AAT. The task goal was to decide if one of the memory letters was present or not in the display string.

So for further example, if the memory-set was 'C,O,V' target present strings could be of differing lengths: -

D C G K N

E O T

S V J L X F A

Each requiring an affirmative response, whilst a negative one should follow a string such as 'D F G K N W R'. So like the AAT a two-alternate forced choice was required. The rest of the methodology closely followed that set out for the AAT by Haider and Frensch (1996). Briefly, participants were told that one of the memory items could appear anywhere in the string, though in fact if a memory-target was included in a display string it always appeared in the same position, second from left in a string. This relegated string items that followed this position redundant to task completion.

Again, as in the AAT, the varying lengths of the display strings allowed for an analysis of the response times to determine whether the participants had learned to ignore the redundant items or not. If they were not able to ignore the redundant letters in the search, then response times would rise and fall systematically with string length so that longer response times would follow longer strings. Alternatively, if participants did learn to ignore the irrelevant letters this string-length effect should not be apparent, or at least should gradually reduce as practice progressed.

As can be seen from this description of the Target Search Task (TST), the set-up was in some ways very similar to the alphabet arithmetic task; however the objective was to produce a conceptually different task. The TST as set-out above did seem to achieve this in a number of ways. Task requirements for the TST were different from the alphabet verification task: the TST required just a search for a target. Having no brackets, just a uniform string to search meant the TST was perceptually different to the AAT. The TST was also a unitary task that did not decompose into the calculation and recognition sub-tasks found in the AAT.

Perhaps most importantly considering the occurrence of mnemonics in study 1 chapter 4, instance learning would be very unlikely. Since the TST was made-up from any random combination of letters, the number of possible unique instances was not constrained in the same way as the earlier strings which had to follow the alphabet. As a result strings did not repeat in the blocks.

## **5.2.2 METHOD**

### **DESIGN**

A three factor mixed design was used - Error Type (Trained-Position vs. Post-Trained-Position – within subjects) \* String Length ( five levels, 3-7 letters – within subjects) \* Blocks ( 1 or 5 – between subjects). The dependant measure was Mean error rate per block and RT.

Two levels of training were used, one block and five block. The one block condition was ran to determine a practice effect was occurring and any measured reduction was not something participants immediately adopted. As in the previous chapter, the error scores of interest were in the test block following training. Particularly, the comparison of error rates for targets at the trained position vs. targets in novel positions.

### **SUBJECTS**

Subjects were 38 students and staff at the Open University Milton Keynes, randomly ascribed to one of the two conditions, 2 or 6 blocks. Their ages ranged between 28 and 64 with a mean of 42. Depending on condition (2 or 6 blocks), participants received either £4 or £7. one subject's data was excluded from the analysis, as their results clearly indicated a misunderstanding of the instructions.

## MATERIALS

### *Stimuli*

Participants verified strings presented in either 1 or 5 training trial blocks. In each block there were 50 trials where the memory target item (C,O or V) was present at the second position from left, and 50 trials where the target item was absent. The string length varied from 3 to 7 items long, with 10 examples of each length generated for both present and absent conditions. This gave the total of 100 trials per block (this was repeated to create the 5 training blocks).

Following the training trials, a test block followed which differed in that 25 of the target present trials were constructed so the memory item appeared in the string further right than during training trials. So, the memory item could appear in position 3, 4, 5, 6 or 7 from left.

For example: -

D G K C K

E H O

S A J L Z F V

The position of the memory item was counterbalanced across trials as far as possible. A set of 10 practice trials was also generated (five target present, 5 target absent), used to familiarise participants with the task.

The set of strings were randomly created using a Visual Basic program with the constraint that C,O or V would be in the second position for 50 iterations of the program and absent for the next 50 loops (modified to produce the test trials). This provided the sets of strings for each block, with the letters selected in as random a method as possible.

Strings were presented at the center of a 17" colour PC monitor, viewed from a normal distance. Letters were approximately 0.4\*0.4 cms in size and 0.4cms apart. So strings ranged from approximately 20 to 54mm in total length. Each string was in black against a gray background.

### *Apparatus*

To display the strings and collect the responses, the sets of strings were entered into an experiment program generated with E-Prime (for technical specifications see Schneider, W., Eschman, A., & Zuccolotto, A., 2002a, or Schneider, et.al, 2002b), running in Windows 98SE at 1024\*768 resolution.

Participants responded using a standard PC keyboard by pressing "Z" if a memory item was present or "M" if the memory items were absent. This was reversed for half the subjects, with "Z" indicating the absence of the target and "M" now indicating the presence of the target.

### PROCEDURE

Participants were seated in front of a microcomputer. Before the task began, they were asked to read an informed consent form (see appendix 1) and to sign it if in agreement. After this, the experiment began with participants reading the task instructions presented on the computer monitor. Subjects were told that their task was to first to memorise three letters (C,O and V). Following this, they would be presented with strings of letters and their next task was to determine if one of the memory items was present or not.



They were informed that only one item at most would ever be present. Participants responded using the computer key board by pressing “Z” if a memory item was present and “M” if they were not (reversed for half the participants). Following this, a short practice session commenced, where participants received 10 strings to scan for the target. If more than one error was made, the practice session repeated until the criterion of 90% accuracy was reached.

Once the practice session was successfully completed, participants were told that the actual trials would begin when the spacebar was pressed. The experimenter left the room while the remaining blocks were completed. Overall, 50 strings with the target present and 50 strings with the target absent were each presented once during each training block. This took approximately 15 minutes for the 2-block condition (1 training block and 1 test block) and 40 minutes for the 6-block condition (5 training blocks and 1 test block). In all, each subject verified 200 strings in the 2-block condition and 600 strings in the 6-block condition. The test block followed on from the training blocks with no indication of the change to the participant.

The order in which the strings were presented was randomly determined for each participant in each block. A trial began with the presentation of a cross fixation point displayed for 500ms, replaced by the string, which remained on the screen until a response was made. When an incorrect response was made, an error message “incorrect” appeared for 500ms accompanied by a beep from the computer speaker. For a correct response only a “correct” message appeared for 500ms before the fixation was again presented.

Participants were allowed to pause between each block if they wished. A maximum pause of 10 minutes was allowed before the next block appeared automatically. Upon completion of the experiment, the experimenter returned and asked set questions about the strings (see appendix 2). This was directed at determining if participants changed their strategy over practice and were aware of the underlying rule fixing the target position. Questions began with general open-ended enquiries about the strings and eventually focused on the in-built regularity with an explicit question. Following this, participants were asked if they had any questions, then they were de-briefed about the purpose of the study, paid and allowed to leave. The 1-training block condition took on average 20 minutes to complete and the 5-training block condition averaged 51.5 minutes

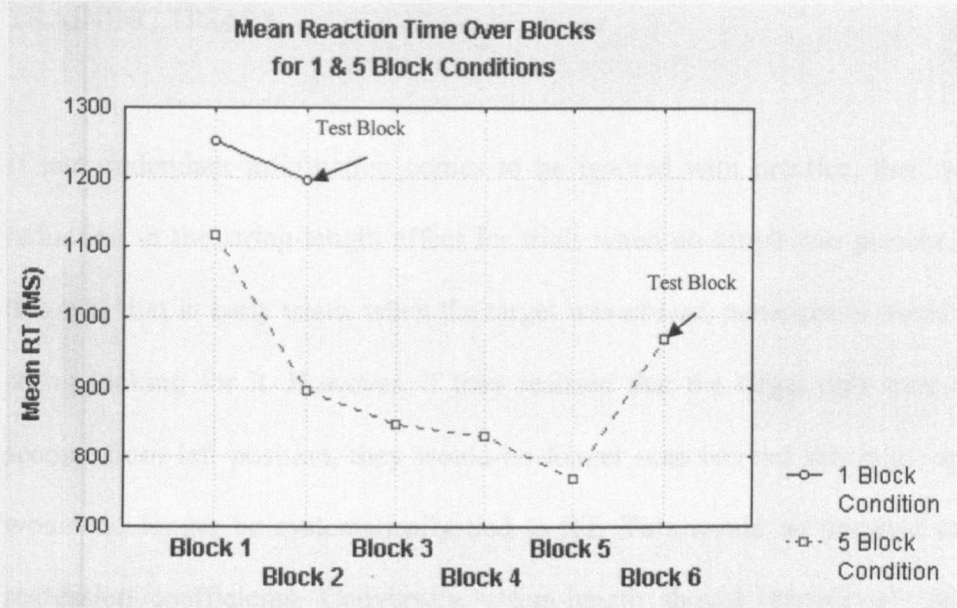
### **5.2.3 RESULTS**

#### **OVERALL ANALYSIS**

To enable the effect of string length on RT to be compared directly, for all experimental conditions the best fitting linear regression lines were computed separately for each subject and each trial block. The regression coefficients obtained were then subjected to ANOVA. The analysis followed the same form as in study 1 in the previous chapter.

Mean error rates were computed for each subject in each practice block, subjects were excluded from the analysis if their error rate was higher than 10% throughout the training blocks ( $N = 1$ ). This resulted in 18 subjects in the 1 training block condition and 19 in the 5 training block condition. Mean error rate across the two conditions ranged from 0 to 8%

There was no speed-accuracy trade-off. The correlation between (RT) and mean percent error rate for the 5 training block group was  $r = -.258$ ,  $p > 0.05$  (excluding responses to strings where the error could occur in the redundant segment) and  $r = -.540$ ,  $p > 0.05$  for the 1 training block condition. Figure 5.1 illustrates the mean RT for responses to all strings for both training conditions.



*Figure 5.1: Mean Response Time (RT) for string verification over blocks for both 2 and 6 block conditions for Experiment 2.*

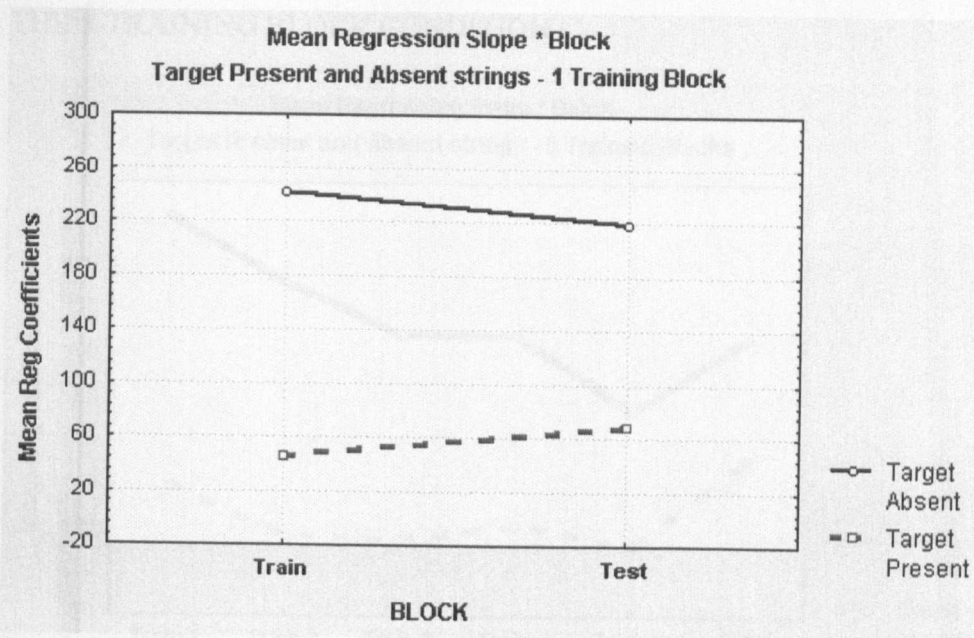
As can be seen, RT appeared to reduce over blocks for both 1 and 5 training block conditions. A One-Way ANOVA (Block 1-2, within subjects) revealed no main effect of Block for the 1-Training Block condition. So RT was similar for both the one training and test block. However, there was a block main effect for the 5-training block condition (Block

1-6, within subjects)  $F(5,90) = 23.56$ ,  $p < 0.001$ . Post hoc tests (Tukey) revealed a significant reduction in RT between the first and last training blocks ( $p < 0.001$ ) and a reliable increase in RT between the last training and test blocks ( $p < 0.001$ ). This showed that with *more* than one block of training, RT can decline in the task, and this trend reversed for the test block.

#### TARGET ABSENT/PRESENT STRINGS: STRING-LENGTH EFFECT DURING TRAINING TRIALS

If task redundant information comes to be ignored with practice, then we should see a reduction in the string-length effect for trials when no target was present. The reason for this was that in early trials, when the target was *absent*, participants would scan the whole string looking for it. However, if they realised that the target only ever appeared at the second from left position, they would no longer scan beyond this point and string-length would no longer be systematically tied to RT. This would be revealed indirectly by the regression coefficients. Conversely, string-length should immediately have little effect when the target was *present*, as participants would stop scanning as soon as it was found in the second position, no matter how long the string.

## THE 1-TRAINING BLOCK CONDITION



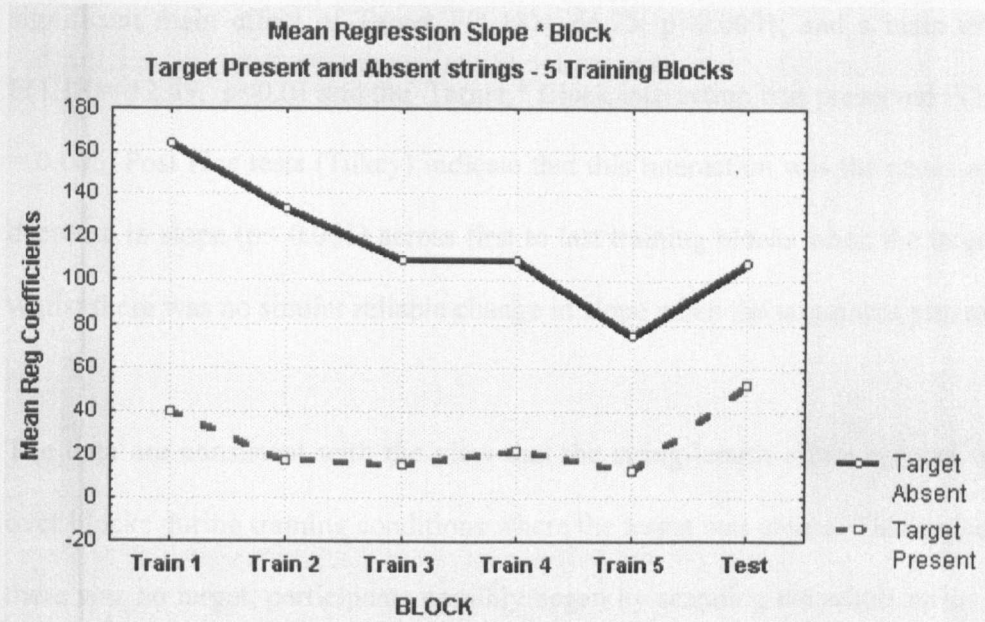
*Figure 5.2: Mean regression coefficients over Training and Test blocks for 1 training block conditions: Target Present and Absent strings, for Experiment 2*

The regression slopes calculated for each subject by block for the 1-training block condition were subjected to a 2-Way ANOVA within-subjects design 2(Target – Absent, Present. ) \* 2(Block – 1,2). As indicated in figure 5.2, a main effect of Target type was revealed  $F(1,17) = 20.50$ ,  $p < 0.001$ , but there was no effect of Block ( $p > 0.1$ ).

This suggests that after one practice block, the string-length effect was significantly greater for Absent than Present trials, as expected. Additionally there was no significant difference between practice and test blocks, indicating IR requires practice to develop. That is, no reduction had occurred after just one block, so the coefficients were similar in both training and test phases. This claim should be supported by the test error data.



## THE 5-TRAINING BLOCK CONDITION



*Figure 5.3: Mean regression coefficients over Training and Test blocks for 5 training block condition for Target Present and Target Absent strings, for Experiment 2*

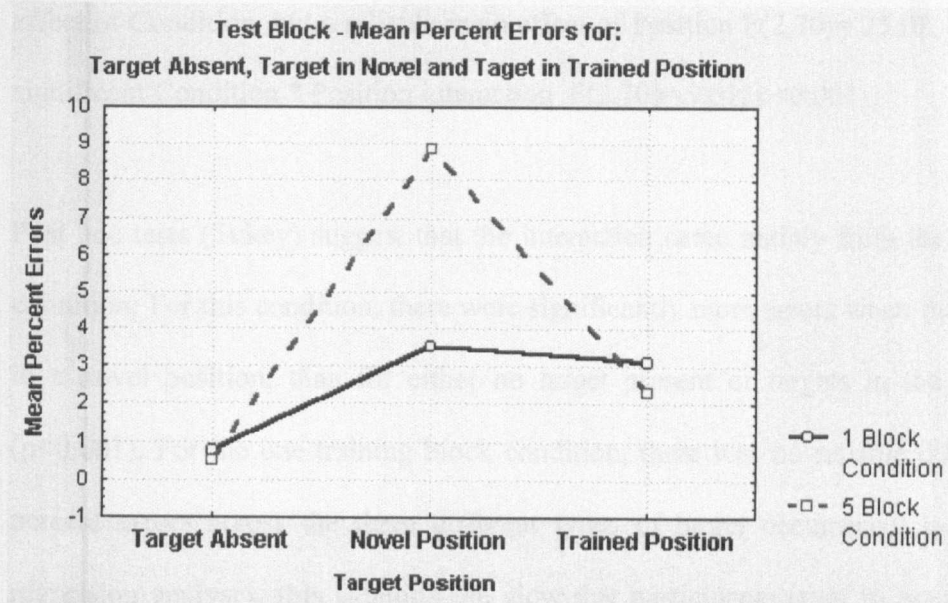
As can be seen from figure 5.3, the regression coefficients start out higher for target absent than target present trials as expected, while the slope seems to reduce over trials for the target absent strings. When the coefficients for both the target present and absent five training blocks were analysed a 2-Way ANOVA 2(Target, Absent & Present – within subjects) \* 5(Block, 1 to 5 – within subjects) revealed a significant main effect of Target type  $F(1,18)= 100.79, p<0.0001$ , and a main effect of Block  $F(4,72)= 6.92, p<0.001$  and a Target \* Block interaction  $F(4,72)= 3.77, p = 0.008$ .

A follow-up analysis comparing first to last training blocks only: 2-Way ANOVA 2(Target, Present & Absent – within subjects) \* 2(Block, First & Last – within subjects), found a significant main effect of Target  $F(1,18)= 66.73$ ,  $p<0.0001$ , and a main effect of Block  $F(1,18)= 12.89$ ,  $p<0.01$  and the Target \* Block interaction was preserved  $F(1,18)= 9.72$ ,  $p = 0.006$ . Post Hoc tests (Tukey) indicate that this interaction was the result of a significant decrease in slope ( $p< 0.001$ ) across first to last training blocks when the target was absent, whilst there was no similar reliable change in slope when the target was present.

The data are consistent with the view that the string-length effect reduced systematically over blocks during training conditions where the target was absent. This suggests that when there was no target, participants possibly began by scanning the whole string, then learned to ignore the irrelevant items. Although (as pointed out by Haider & Frensch, 1996, Expt. 2) as for the AAT, an alternate explanation could be faster scanning of this irrelevant segment. A decrease in slopes over blocks was not found or expected for trials when the target was present. As there was always only one target, participants would respond when finding this and so varying string lengths would have little effect.

## ERROR RATES

The error rates across the test block were analysed to determine the effect of moving the target letter from its constant second position in the practice trials to a position outside this in the test block. If participants came to rely on the position of the target over practice and subsequently did not check the remainder of the string, more errors should occur in the test block when this strategy fails for targets in a novel position.



*Figure 5.4: Mean percent error rate for the final test blocks where target could appear in the trained or a novel position; target absent string errors added for comparison, Experiment 2.*

As can be seen in Figure 5.4, there seem to be a greater percentage of errors in the test block following five as compared to one training block. This was when the target occurred in a novel position compared to when either there was no target or a target was in the trained position. Following one training block, at test there appeared to be more errors when the target was present compared to absent. However, there was little differentiation caused by target position in this case.

A 2(Condition – 1-training block & 5-training blocks, between subjects) \* 3(Position – No Target, Target Novel Position & Target Trained Position) ANOVA revealed no significant



effect of Condition, but a reliable main effect of Position  $F(2,70)= 25.02$ ,  $p< 0.0001$ , and a significant Condition \* Position interaction  $F(2,70)= 9.01$ ,  $p<0.001$ .

Post hoc tests (Tukey) suggest that the interaction came mainly from the 5-training block condition: For this condition, there were significantly more errors when the target appeared in a novel position, than for either no target present or targets in the trained position ( $p<0.001$ ). For the one training block condition, there was no reliable difference in mean percent errors across the three different types of target occurrence. In addition to the regression analyses, this supports the view that participants learn to ignore the irrelevant string items and not just process them more quickly.

It would also seem that practice with more than one hundred trials was required for this to occur. Again this concurs with the regression analysis for the 1-training block condition. Overall, this indicates a practice effect was occurring in the TST task, and part of this change was due to a reduction in processing, as suggested by Haider and Frensch's (1996, 1999a, 1999b) IR hypothesis.

#### INTRODUCTION OF ERRORS IN THE POST-TRIPLET STRING: RESPONSE TIMES (RT).

For the final test block, following longer practice participants seem to make more errors when targets appeared in an untrained novel position. Suggesting they were not processing all the string at this point. However, since feedback was given after each trial, it could be expected that the subsequent increase in error feedback would prompt scanning of the whole string once again.

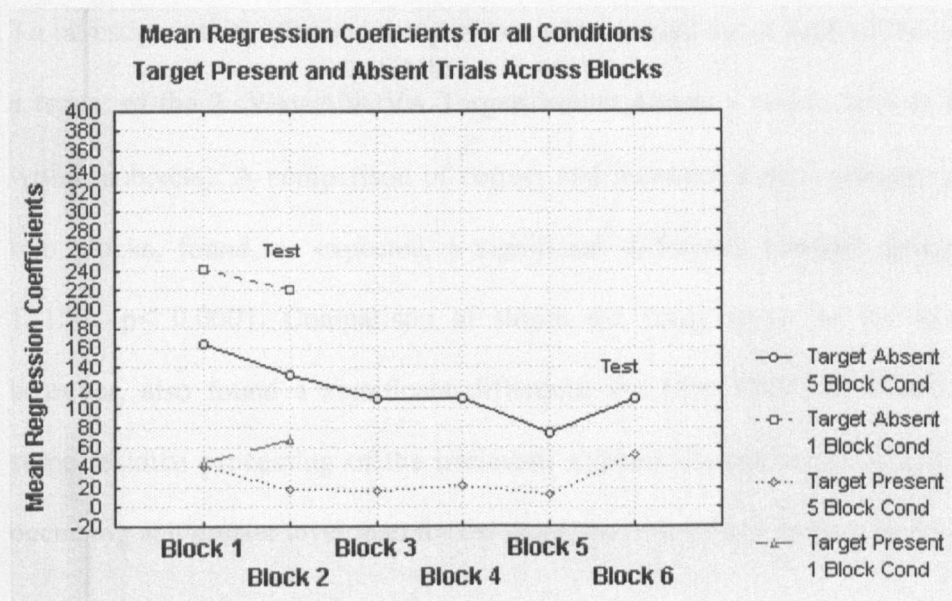
If this was the case there should be a re-instatement of the string-length effect for target absent strings, shown by an increase in the regression coefficients. That is, the string-length should once again predict RT. As reported above, there was no reliable change in coefficients between train and test for the 1-block condition (Fig. 5.2), Figure 5.3 however displays a difference between these blocks following five hundred training trials.

The change in slope shown between the last training and test blocks for target absent strings in figure 5.3. for the 5-training block condition, was subjected to a 2(Target – Present & Absent, within subjects) \* 2(Block – Last Training & Test, within subjects) ANOVA. There was a significant main effect of Target  $F(1,18)= 22.22$ ,  $p< 0.001$  and Block  $F(1,18)= 14.71$ ,  $p< 0.01$ , but no Target \* Block Type interaction.

This suggests that the scan length for the strings did increase when participants realised they were making errors, but this was true for both target present and absent strings. This probably reflects some confusion when response errors start to unexpectedly occur: slopes for target present are still reliably lower than for target absent at test (Tukey,  $p<0.05$ ). Overall it does seem that scanning of the formerly irrelevant items did re-emerge in the test block, in a similar fashion to that found with the AAT in Study 1. This further supports the view that IR did occur in the TST, but reduction lessened when prompted by error feedback.

## RESIDUAL PROCESSING OF STRINGS

The above analyses using the methods laid out by Haider and Frensch (1996) suggest that within this target search task, IR does occur over practice. This was supported by both the reduction in the mean regression slope for the 5-training block condition (target absent), and the increase in errors for the same condition when the target occurs outside position two. However, as can be seen from figure 5.5, the results are not as convincing as for the AAT in Study 1. Even at five training blocks there appears to be a difference between the coefficients for present and absent trials.



*Figure 5.5: Mean regression slopes for all conditions, target absent & target present trials, Experiment 2*

Earlier analysis of the 1-training block condition indicated a significant difference between target present and target absent trials at both training and test as expected. The hypothesis was though, that after the additional training in the 5-training block condition, this difference should reduce. So after five training blocks, increased IR should mean both present and absent coefficients should approach equivalence. That is, towards the end of practice, coefficients should be close to zero for both target present strings and target absent strings. So for the longer training condition there were two points of interest, the initial difference between string types and the difference after training was completed. A difference was expected for the former but not the latter.

To investigate this, planned comparisons were carried out at each of these times following a repeat of the 2 -Way ANOVA Target(Present,Absent – within subjects)\*Block(1 to 5 – within subjects). A comparison of correct and incorrect slopes, collapsing across the first two blocks, found, as expected, a significant difference between string types  $F(1,18)=131.84$ ,  $p<0.0001$ . Comparison of slopes for string types for the last training block however, also found a significant difference  $F(1,18)=15.07$ ,  $p=0.001$ , suggesting that some residual processing of the irrelevant segment for the target absent strings was still occurring at a greater level than for the target present strings, even after 500 training trials

The more stringent Post-hoc test (Tukey) with multiple comparisons of all the blocks in the ANOVA, still indicated that for the last training block there was a significant difference between the mean slopes for target Present vs. Absent ( $p<0.001$ ). This difference is important, since the mean slope for target present strings can be taken as a baseline. One which represents the complete adoption of the reduction strategy. This is because when a

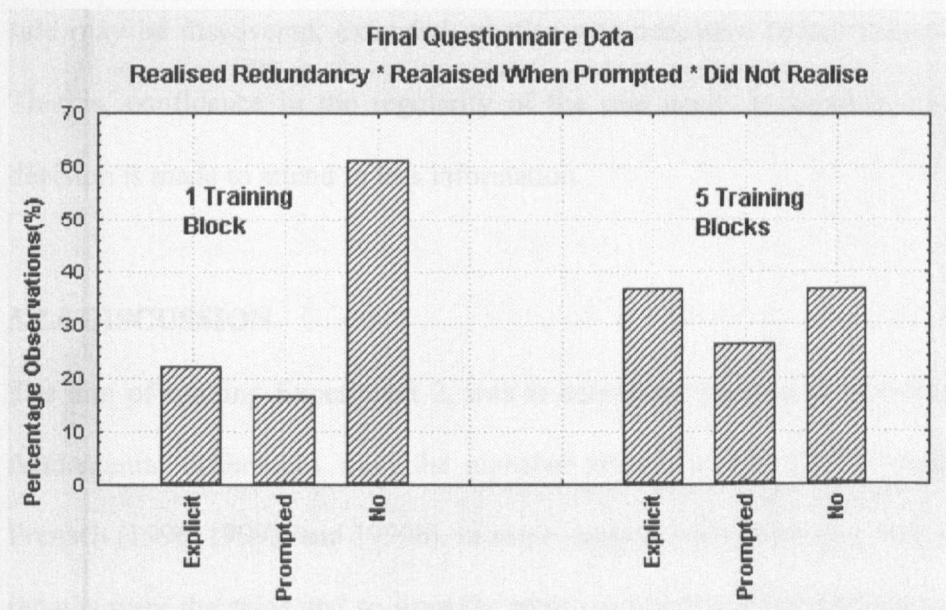
target was present, the string search terminated when it was reached. Thus, this would be equivalent to the constrained search following *complete* reduction when the target was absent.

However, the analysis show that at the end of the five training blocks, there was still a significant difference between string types. While the target absent slope did reduce over training it remained reliably different from the target present equivalent. This suggests that there was still some residual processing of the redundant part of the string even at the end of practice block five. This residual processing may just occur as a normal part of scanning behaviour or participants may be expecting some trick trials. The anticipation of trick trials was not really indicated in the de-briefing sessions. This will be revisited in the discussion.

## END OF SESSION QUESTIONNAIRES

Following the completion of the string verification trials, each participant was again asked a number of questions (see appendix 2). As before, the questions began as a general enquiry over whether anything unusual was noticed about the strings. The questions became gradually more specific about the position of the memory-targets. The aim was two fold. First to determine if the underlying rule and redundancy was consciously noticed and second, to find out if any particular strategies other than IR had been employed. Figure 5.6 shows for each condition the percentage number of participants that explicitly stated knowledge of the redundancy without prompting, those who realised it when they were prompted by the suggestion of the redundancy and those who had not realised the rule at any level.

For the 1-training block condition it seems larger percentage did not realise the rule than did. This was to be expected after just one training block. For the 5-training block condition the number who knew the rule was approximately equivalent to those who did not. Over 25% also realised the rule when prompted. Taken together this suggests more had some idea of the regularity than did not after five blocks (accepting that prompting can lead to affirmative demand characteristics).



*Figure 5.6: Participant responses to final questionnaire: Those that explicitly realised post-triplet redundancy, those that realised it when prompted and that did not know of the rule. Presented as a percentage to allow for different group sizes, Experiment 2.*

Comparing across conditions, it seems that overall more people appear to learn the regularity of target positioning following longer practice than short, both explicitly and under a prompt. Interestingly, over 60% and 35% did not realise the rule at any level for short or long training respectively. This indicates that learning the regularity in this task

was not that trivial. In the previous thesis study using the AAT, only 1 participant in each condition did not realise the rule after prompting. It is possible that the task may not be difficult enough to encourage individuals to look for a way of reducing effort.

Overall, it does seem that some participants learn the regularity within 100 trials, though more learn it following further training. However, the regression and error analyses indicated that no reduction occurred after just one training block. Thus, it seems while the rule may be discovered, extended practice was necessary before reduction was adopted. That is, confidence in the regularity of the rule needs to develop, before a conscious decision is made to attend to less information.

#### **5.2.4 DISCUSSION**

The aim of running Experiment 2, was to determine whether IR occurred for a task with fundamental differences from the alphabet arithmetic task (AAT) used by Haider and Frensch (1996, 1999a, and 1999b). In some respects both tasks were very similar, the string lengths were the same and so were the trials per block and required response. Importantly, both tasks had an underlying rule, which when selected relegated one part of the stimulus irrelevant to correct task completion. However, the two tasks did also differ in a number of dimensions.

#### **COMPARISON OF AAT AND TST**

##### *String segmentation*

First, in the AAT, both the task instructions and the visual properties of the strings served to divide each into two distinct sources of information; triplet and post-triplet. Even in

variations on the task, such as those used by Lincourt et.al (1997) had this attribute. There was more than one triplet, but the string was still broken up in two or more segments by both the task instruction and appearance. In the Target Search Task (TST) used in the present experiment, neither the visual appearance of the string, or the task instruction guide participants to form separate information sources. Thus, the question was whether IR still occurred when relevant and irrelevant parts of the string did not fall into pre-defined separate sources of information. That is, were participants able to segment the string into relevant and irrelevant sources themselves, and then ignore the irrelevant segment?

### *String learning*

Second, in the AAT, focusing attention on the triplet through the calculation necessary to verify it, may serve to make this segment memorable. As there were only 10 unique correct and 10 incorrect triplets used, each repeated 10 times in a block. It may become possible to recognise correct and incorrect triplets as practice proceeds. Indeed experience of using the AAT suggested this was the case. A number of participants reported the formation of mnemonics to aid string verification and avoid any triplet calculation. Since IR is reliant on the detection of an inbuilt regularity in the stimulus (only one part of the string is always relevant), realising that the triplets were repeating may be necessary to accept this regularity and adopt reduction. The TST used in the current study did not split the string up into memorable segments and anyway each string was unique so this type of learning could not occur.

Of course, memorising the triplets could not completely explain the IR effect. Haider and Frensch (1996) found that the reduction in the string-length effect transferred to novel



stimulus sets, ruling out an instance based account. However, even though the string-length effect did not significantly increase, they did find an increase in RT on switching stimulus sets. This does suggest the possibility that some stimulus-specific aspect of the strings had been learned. It is still possible that early recognition of strings may increase confidence in the formation and reliance on the IR strategy. That is, when participants started to recognise triplets repeating in the AAT, they were likely to be more confident in relying on the triplet as a flag indicating the correct response. So allowing the development of IR. The necessity of such recognition could be determined by the current paradigm where such learning was unlikely to have occurred.

#### EVIDENCE FOR INFORMATION REDUCTION

The results from Experiment 2 suggest that IR occurred within the TST. The mean regression slopes for the target absent trials reduced over training. This indicated that reaction time was no longer systematically tied to string length. This indirectly suggested that participants were focusing attention mostly at the second item position from left. Additional support for this was found when the error rate was investigated. Significantly more errors were made in the test block when the target appeared in the formerly irrelevant segment of the string (i.e. outside position 2). Again this sustains the view that participants had learned not to attend to this part of the string. So the target was missed when it appeared in the novel position.

#### GENERALISING THE IR EFFECT

Overall then, the results indicate that IR also occurred for the TST used here. This result is interesting, particularly as it generalises IR beyond just the AAT. Since Haider and Frensch

(1996, 1999a, 1999b) propose that IR is a general phenomenon that occurs following practice at a task, such generalisation is important. To accept that IR needs to be incorporated in contemporary theories of skill acquisition, such as Logan's (1988) instance theory, Anderson's (1998) procedural model and Newell's (1990) theory of cognitive architecture, IR must be available across many task domains and not be unique to very particular circumstances. Accordingly, it is encouraging to find reduction in information processing for a different type of task.

#### *IR and Logan's (1988) instance theory*

What this study seems to support is the proposal that IR is an aspect of skill acquisition that is outside that described by Logan (1988, 1994, 1996, 1999). Logan's main assertion is that the improvement that follows practice is the result of a switch from an algorithmic method to the direct retrieval of a solution from memory. The switch occurs when sufficient instances of the situational solution have been stored, so that they are retrieved more quickly than the algorithmic route. So skilled performance reflects the direct speeded access to a memory that solves the problem.

This theory is much cited and explains a good deal of empirical research. While IR is not stimulus specific (Haider & Frensch, 1996), some form of instance learning did appear to take place in the AAT: certainly there was report of the use of mnemonics to avoid the algorithmic calculation of triplets (Study 1, chapter 4). As already intimated, it could be that some form of instance learning moderates IR, which would be problematic when excluding Logan's (1998) account. However, such instance learning was unavailable for the TST, as each string was unique.

Since IR occurred also within this task it can now be concluded that instance learning, as strictly defined by Logan (1988), is not seemingly necessary for either the development or employment of such a reduction in processing. This is not to say that the individual instances are not encoded or even that some abstraction across them may occur. As suggested in the literature review, how the task is achieved in each trial may be ‘remembered’ and this may guide the method used as practice progresses. This is not, though, really the type of learning that is central to Logan’s instance theory.

#### *IR and separate information sources*

At a less theoretical level, the TST differs from the AAT in that the former is a unitary task while the latter splits into two sub-tasks. That is, the alphabetic arithmetic strings were made-up of the triplet calculation and the following over-learned alphabet recognition. It was supposed that maybe some division of relevant and irrelevant information at a perceptual or cognitive level was necessary for IR to progress over practice. Since the AAT, in one form or another, was repeatedly used this had not really been tested (accepting Lincourt et al. (1997) who showed triplet salience was not a factor, though not the effect of separate information sources). This was an important aspect to investigate, since relevant and irrelevant sources may not always be so easily distinguished..

In vision research there is evidence that stimulus properties that serve to separate target from distracters, improve task performance (Olson and Chun, 2001, Snowden & Edmunds, 1999, Ball & Sekuler, 1981), and this improvement is probably the result of selective allocation of attention at some level. So it is possible that some form of pre-existing task

segmentation could promote IR. The current study using the TST makes this unlikely, as IR seemed to occur even though the stimulus strings were unitary and did not split into sub-tasks at either a perceptual or cognitive stage. The strings contained no different character, like brackets at the relevant segment and there was one task that applied overall: to search the whole string for target letters. Using a methodology very similar to Haider and Frensch (1996) their criteria for IR were met: therefore, pre-defined relevant and irrelevant areas are not necessary for reduction.

## SUMMARY

The TST results indicate that IR has now been shown to generalise across at least two different tasks. Further, recognition of repeating strings is not necessary for it to occur. This rules out a strict instance account as an explanation of how IR is either learned or deployed. It also appears that information sources that are perceptually different, or split into sub-tasks, are not required for the regularity to be learned and reduction to take place. The results extend the conditions under which IR can develop.

### *IR in the AAT and TST: Error rates*

However, if Experiments 1 and 2 of the thesis are compared, the results are not quite so convincing for the TST. The percentage response-errors during the test phase when targets occurred outside the trained position were much less for the current task: 9% for the TST, but 43% for the AAT in the longer training conditions. For this to arise, scanning of the whole string at test must have resumed more quickly for the TST.

One explanation for this is that for the AAT, when error feedback occurred during the test phase, participants could ascribe the error to either a miscalculation of the triplet or accept the much less likely case that they missed a post-triplet error. Assuming a post-triplet error was less likely since training experience suggested that errors did not occur there. In comparison, for the TST, error feedback at test indicated only that a target had been missed; no other reason could be ascribed to the error. So, participants would perhaps search more carefully and quickly include the whole string.

#### *IR in the AAT and TST: Regression coefficients*

More problematic is the degree to which IR ensues in the different tasks. There appears to be more residual processing after practice in the TST compared to the AAT. This can be seen comparing the mean coefficients of both target Present and target Absent trials at the last training block. It appears participants were to some extent, still scanning beyond position two in the string when the target did not appear at this position. In comparison, careful examination of the Haider and Frensch (1996, 1999 & 1999) research and the findings in our own study using the AAT, suggest that this residual processing did not occur to the same extent for the original task. In thesis Study 1, by 500 trials IR seemed to have progressed to the point that the string-length effect was approaching zero. Haider and Frensch (1999a, pp131.) accept that IR has generally happened by five hundred trials and examination of their research supports this. Thus, it appears as though the original AAT and the current TST differed in at least one critical way, moderating IR. This raises the question as to what the key difference(s) might be.

### *Task difficulty*

Separate information sources and the generation of mnemonics have already been mentioned, and these differences across tasks may have caused some moderation of the effect. Though, perhaps the most parsimonious explanation focuses on the varying levels of difficulty associated with the different tasks. The Alphabet Arithmetic task was arguably more difficult than the Target Search task. In Study 1 (using the AAT), each string took longer to verify than for the TST used currently. So, it maybe that the search task was so simple to complete, that the motivation to reduce was less. Certainly, this would make sense, why reduce the information processed when it can be coped with so easily. One could alternately suppose that maybe participants just did not believe the target would always occur at the same position. However, this was not suggested by any of the participants during the de-brief session.

One way to test if task difficulty is playing some role in moderating IR, is to increase the cost of scanning the whole string and the overall task difficulty. The third experiment was aimed at achieving this by increasing the length of the strings. Scanning the whole instance then becomes more effortful, increasing the possible benefits of an IR strategy.

### *Conclusion*

From a common sense viewpoint it is likely that IR plays an important part as one component in skill acquisition during daily life. IR within the TST suggests that when a field of targets were searched, this search may become constrained to likely target locations. Assuming experience that suggests such a regularity. As proposed in the introduction for the present study, this is likely to have further ‘real world’ implications.

As reported from the recent army science conference (Mouloua *et. al.*, 2004, Diaz *et. al.*, 2004), available information in military and aviation theaters is ever increasing in volume and complexity. Therefore, searching for target information becomes more problematic and error prone. If IR is engaged in speeded high pressure environments to cope with this load, then understanding the task attributes, such as task difficulty, that lead to and moderate the strategy is worthwhile.

### **5.3 EXPERIMENT 3: INFORMATION REDUCTION IN A TARGET SEARCH TASK (TST): LONGER SEARCH**

*The aim for this experiment was to determine if increasing the length of the stimulus strings moderated information reduction, and specifically reduced the level of residual processing of the task-irrelevant items*

#### **5.3.1 INTRODUCTION**

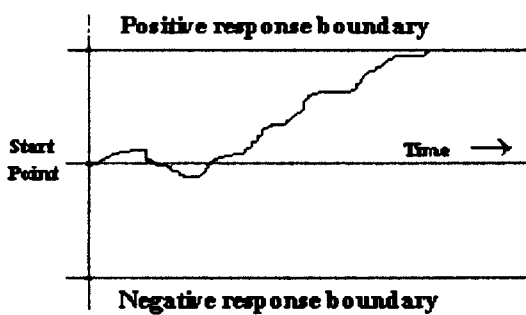
In the preceding study of the thesis, information reduction (IR) was shown to also occur for a new task, where the aim was to search for items from a memory-set in an array of distracters. This encourages the view that IR can arise in different tasks and consequently may be a regular part of skilled performance. However, the effect did not appear quite as convincing as when the AAT was employed in thesis study 1 and the earlier Haider and Frensch (1996, 1999a, 1999b) research. Perhaps the most economical reason to explain this was that the task was just not difficult enough for IR to become worthwhile. Certainly, IR appears to be moderated by task dimensions, such as instructions to be speeded or accurate while verifying a string (Haider & Frensch, 1999a). So, IR it seems is more likely when there is some time pressure on processing the strings; task difficulty may also have such a moderating effect.

#### *Random walk models*

That task instruction moderates IR (Haider & Frensch, 1999a) is particularly interesting. Not only does it suggest that IR was a conscious decision, but also indicates one of the conditions under which it is most likely to occur. To explain the different degrees of IR found under speed vs. accuracy instructions, Haider and Frensch (1999a) refer to random



walk models, such as Ratcliff's (1981) diffusion model. The underlying premise of the model is that information is accumulated gradually until either a positive or negative response is made. Further, the criteria determining how much information is acquired before a response, is a variable controlled by the individual (Ratcliff, 1985). In the typical sequential letter-matching task used by Ratcliff (1981), the criteria can be thought of as upper and lower boundaries.



*Figure 5.7: Random Walk Model illustrating upper and lower response criteria*

When the upper boundary for information is reached an affirmative response occurs, while a negative response ensues when the lower boundary is reached (refer to figure 5.7 for clarification of the boundary positions). One factor that influences the setting of these boundaries by the individual may be task instructions: under accuracy conditions the boundaries would be set far apart, so greater information would be accumulated before a response is triggered. Conversely, under speeded conditions, the boundaries would be set closer together, so less relatively information would be sought before a response.

A similar process was suggested for IR under different task instructions (Haider & Frensch, 1999a). When instructed to be accurate, response criteria are set further apart, so much of the available information is accumulated before string verification takes place, so there is little IR. The reverse is true under instructions to be as quick as possible: criteria are set closer together so a response is made with a reduced information search, so IR increases. This is not just a speed-accuracy trade-off in their data, as increased errors during the test phase are specific to the formerly irrelevant items. As Haider and Frensch (1999a) suggest, under a rational view, the setting of more liberal criteria within a random walk process is likely to be accompanied by a similar reduction in the processing of irrelevant string items. Their data suggests this is even more pronounced when not only the instructions encourage speed, but also the stimulus duration is progressively reduced forcing a quick response.

This would indeed seem to explain the empirical data (Haider & Frensch, 1999a). Important for the current argument though, is that Ratcliff (1985), also proposes some other factors that may shift the response criteria. He reports a number of studies to show that both probability and task difficulty seem to change the response criteria. So in the language of random walk theory, the TST may have such a low level of difficulty that the response criteria are set wide, subsequently under these stricter criteria IR does not progress so well. Therefore, under a similar argument to Haider and Frensch (1999a), task difficulty may also moderate IR; increasing difficulty in the TST should then increase the IR effect.

### *Selective attention*

As suggested in the literature review it is possible to categorise IR as an attentional 'strategy'. Attention is withdrawn from the irrelevant items and focused on the relevant

segment of the strings. This draws very similar parallels to much of the research on attention. Consequently, this research may provide some clues about the mechanism of IR. Many 'attentional' tasks investigate the effect of irrelevant items on task performance during goal-directed behaviour that focuses on goal-relevant targets (e.g., Bundesen, 1990; Duncan, 1980; Eriksen & Eriksen, 1974; Eriksen & St.James, 1986; Treisman & Geffen, 1967; Treisman, 1969).

A distinction is required at this point though, that differentiates this type of research from IR. Whilst it is true that in these attentional paradigms participants focus on target areas whilst ignoring irrelevant distracter stimuli, they by and large are instructed as to what the targets are. That is, they are informed where to target attention and what to ignore. So, while IR seems very similar to paradigms such as flanker tasks (Erikson & Erikson, 1974), tasks investigating IR are different, in that participants are not informed of the target region. Rather they are led to believe the whole stimulus is relevant. IR arises when the underlying regularity is realised incidentally and used to reduce processing. Accepting this caveat, attention research may prove interesting, especially for an investigation of irrelevant stimuli and task demands.

Of particular interest is the load theory of selective attention (Lavi, 1995, 2000; Lavi, Hirst, De Fockert & Viding, 2004). This theory was developed to account for discrepant evidence that supports both an early and late selection view of attention. That is, there is evidence that suggests selective attention may occur at an early perceptual stage, from this perspective distracters are not processed at any higher cognitive level. However, there are also findings that suggest selection can happen at a later cognitive stage, allowing for some

processing of unattended stimuli (Lavi *et al.*, 2004). These stimuli are then *actively* ignored before a response is made. Lavi (1995, 2000) proposes that the discrepancy can be resolved with a hybrid model of attention. A model that makes the distinction between two types of processing load: perceptual and cognitive or working memory (WM) load.

For tasks with high perceptual load, the theory is that as perceptual resources are used up distracters are less and less likely to be perceived and affect task performance. This accounts for the ‘early’ selection results. Under low perceptual load however, spare resources ‘spill over’ and distracters become perceived, though normal individuals can still select the relevant stimuli to control responses. This is the late selection account, one that requires some active control process to explain behaviour. So processing may be ‘early’ or ‘late’ depending on perceptual load. Further to this though, Lavi *et al.* (2004) indicate that selective attention at the later cognitive stage may additionally be affected by cognitive load. Since at this later stage of processing, some active control over processing is required, increasing task (WM) load will use up resources and hinder this active control. Under these conditions irrelevant stimuli may become difficult to ignore and impinge on task performance. So while perceptual load predicts whether selection is early or late, late selection is further modified by task difficulty.

Abstracting this to the IR hypothesis, once individuals have decided to reduce, irrelevant items may be processed less as perceptual load increases. This is interesting as Haider and Frensch’s (1999b) experiment indicates IR takes place at an early perceptual level. Conversely, it is also possible that irrelevant items may be processed more as WM load

increases. Understanding task difficulty in terms of either perceptual or cognitive load could therefore be an important consideration if such a task attribute is investigated.

### *Strategy selection and task difficulty*

Task difficulty also appears to play a part in strategy selection. If IR is regarded as a strategy then this line of research may be worth considering further. Roberts et al. (1997) and Newton and Roberts (2000) ran a series of experiments in which a *compass point direction task* was used to investigate strategy development. In the task people were given sets of compass point directions, such as one step north, then one step west, then one step south, then one step south, then one step west. The aim of the task was to determine the final co-ordinates relative to the starting point when all the steps had been taken. Their data suggests two strategies seemed to develop in this task: spatial and cancellation methods.

The spatial method generates a full mental representation of the path taken as the steps were completed. So participants using this method would then try to follow the directions internally and form a spatial map of the route they have taken. The cancellation method cancels all the opposite directions and what remains is the answer. So for example if told to take one step north, one step west and then one step south, the north and south steps cancel each other out and you finish one step west of your starting point. (Newton & Roberts, 2000).

The spatial method is slow and error prone, while the cancellation strategy is fast and efficient. People use one of the two methods, but the results are rather counter-intuitive. People with good spatial ability seemed to develop the cancellation method over practice,

while those low in spatial ability tended to stick with the spatial strategy, for which they had less aptitude. The explanation given (Newton & Roberts, 2000; Roberts, 1997) was that having a good spatial representation of the task allowed the regularities to be spotted and the cancellation method to develop. This insight therefore, only becomes available to those with a good representation of the route – those high in spatial ability.

Interestingly, those scoring high in spatial ability did not however develop the cancellation strategy to the same extent if they were provided with a pencil and paper to work out the route (Roberts, 1997). This could be either because the pencil and paper suppressed discovery of the cancellation strategy, or just suppressed its usage. Newton and Roberts (2000) investigated this further and narrowed the reason to a suppression of discovery of the new better strategy. This result could be interpreted to propose new strategies only develop when there is some need to either improve performance or reduce effort. While it is an empirical question if this generalises to other tasks, it does seem logical that strategies develop to meet a goal and may not do so if the goal is already easily attained. To bring this back once again to IR, it would seem possible that IR will increase in the TST if its benefits are more pronounced. The cost of not reducing will become greater as the effort in processing the whole stimulus increases.

In summary, research suggests that increasing task difficulty may change decision criteria so that less information is accumulated before a response is initiated (Ratcliff, 1981). This change is likely to be accompanied by increased IR (Haider & Frensch, 1999a). Increasing perceptual load also makes it less likely that irrelevant information will be processed. Additionally, it would seem that new strategies may only be developed when there is some

significant benefit to do so. Overall then, increasing task difficulty in the TST should be investigated to determine if like task instructions it can moderate IR.

There are a number of ways to achieve this. The TST is quite similar to the Sternberg (1966) paradigm which also uses a memory-set and a target display. The most common way to increase difficulty in this type of task was to increase the number of items for the memory-set. As Lavi *et al.* (2004) suggest, this increases the WM component of the task. While this would make the TST more difficult, increasing the number of memory-set items does not really change the cost/benefit from reducing the number of display items processed. Perhaps a better alternative increases the number of display items, that is, increases the display string lengths. Not only does this increase perceptual load under Lavi and De Fockert's (2003) criteria, but also increases the benefit in reducing, as then significantly fewer items need to be scanned. For that reason, this latter change was realised for thesis Experiment 3. The same method was used as for the previous experiment reported here, except that the number of items in a string now ranged from 5 to 13 letters

### **5.3.2 METHOD**

#### **DESIGN**

A three factor mixed design was used - Error Type (Trained-Position vs. Post-Trained-Pos – within subjects) \* String Length (five levels, 5,7,9,11 & 13 items – within subjects) \* Blocks ( 2 or 6 – between subjects). The dependant measure was Mean error rate per block and RT. The 1-training block condition was again run to ensure IR occurred only after extended practice.

## SUBJECTS

Subjects were 24 students and staff at the Open University Milton Keynes, randomly ascribed to one of the two conditions, 2 or 6 blocks. Their ages ranged between 25 and 55 with a mean of 40. For participation, each was paid £5 per hour, so dependent on condition (2 or 6 blocks), participants received £4 or £7.

## MATERIALS

### *Stimulus*

As for Experiment 2, participants verified either 1 or 5 training trial blocks, in each block there were 50 trials where the memory target item (C, O or V) was present at the second position from left, and 50 trials where the target item was absent. The string length were 5, 7, 9, 11 & 13 items long, for example: -

S C Z Q F

J O K N E T L

A V R L N K Z J U

B C N G H W P K Z Q D

U O W B F K G T S H M Q N

10 examples of each length were generated for both present and absent conditions to give the total of 100 trials a block (this was repeated to create the 5 training blocks).

Following the training trials, a test block followed which differed in that 25 of the target present trials were constructed so the memory item appeared in the string further right than during training trials. So could appear in position 3 to 13 from left. The position was



counterbalanced across trial as far as possible. A set of 10 practice trials were also generated (five target present, 5 target absent), used to familiarise participants with the task. The strings themselves were randomly created using a Visual Basic program as in the previous experiment

Letters in the strings were approximately 0.4\*0.4 cms in size and 0.4cms apart. String length now ranged from 35 to 104mm for 5 to 13 items respectively. Each string was in black against a gray background.

### *Apparatus*

To display the strings and collect the responses, the sets of strings were entered into an experiment program generated with E-Prime (for technical specifications see Schneider, W., Eschman, A., & Zuccolotto, A., 2002a, or Schneider, et.al, 2002b), running in Windows 98SE at 1024\*768 resolution.

Strings were presented at the centre of a 17" colour PC monitor. Participants responded using a standard PC keyboard by pressing "Z" if a memory item was present or "M" if the memory items were absent. This was reversed for half the subjects, with "Z" indicating the absence of the target and "M" now indicating the presence of the target.

### PROCEDURE

The procedure was identical to that used in Experiment 2. As before following a short practice, participants verified either one or five training blocks and one test block. As before, participants were unaware of the underlying regularity and were told the targets

could appear anywhere in the string. The test block followed on seamlessly from training without any indication to participants. As before feedback followed after each trial.

Participants were allowed to pause after each block if they wished. On returning the experimenter asked set questions about the strings as usual (see appendix 2). Then participants were de-briefed about the purpose of the study, paid and allowed to leave. The 1-training block condition took 20 minutes on average to complete and the 5-training block condition averaged 59 minutes.

### **5.3.3 RESULTS**

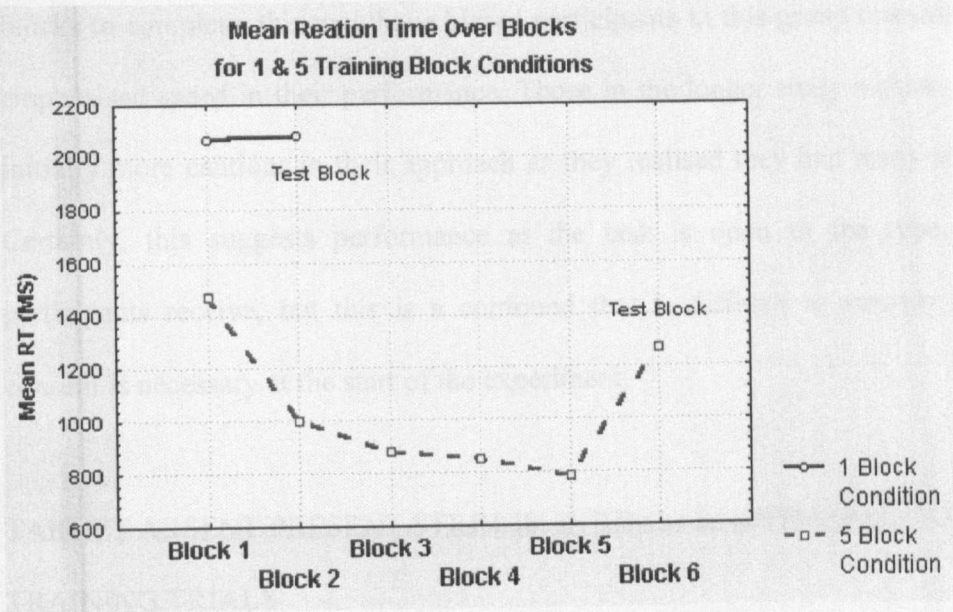
#### **OVERALL ANALYSIS**

To enable the effect of string length on RT to be compared directly, for all experimental conditions as before the best fitting linear regression lines were computed separately for each subject and each trial block. The regression coefficients obtained were then subjected to ANOVA. The analysis followed the same form as in Experiments 1 and 2.

Mean error rates were computed for each subject in each practice block, no participants were excluded as all error rates were below 10% throughout the training blocks. This resulted in 12 subjects in the 1-training block condition and 12 in the 5-training block condition. Mean error rate for the two conditions ranged from 0 to 7% across participants

There was no speed-accuracy trade-off. The correlation between (RT) and mean percent error rate for the 5-training block group was  $r = -.078$ ,  $p > 0.05$  (excluding responses to strings where the error could occur in the redundant segment) and  $r = -.134$ ,  $p > 0.05$  for the

1-training block condition. Figure 5.8 illustrates the mean RT for responses to all strings for 1 and 5 training block conditions.



*Figure 5.8: Mean Response Time (RT) for string verification over blocks for both 1 and 5 training block conditions, Experiment 3.*

As can be seen, RT reduced over blocks for the 5-training block condition, One-Way ANOVA (Block 1-2, within subjects) revealed no main effect of Block for the 1-training Block condition. However, there was a block main effect for the 5-training block condition (Block 1-6, within subjects)  $F(5,55)= 25.77, p<0.001$ . Post hoc tests (Tukey) revealed for this condition a significant reduction in RT between the first and last training blocks ( $p<0.001$ ) and a reliable increase in RT between the last training and test blocks( $p<0.001$ ).

The large difference in RT between groups during the first block is interesting. One possible reason was that each group at the start of the experiment was informed how many blocks they would verify and also the approximate length of the experiment. With only two blocks to complete, this may have biased participants in this group towards a strategy that emphasised speed in their performance. Those in the longer study session may have been initially more cautious in their approach as they realised they had many trials to practice. Certainly, this suggests performance at the task is open to the type of instructions participants receive, but this is a confound that is difficult to remove when informed consent is necessary at the start of the experiment.

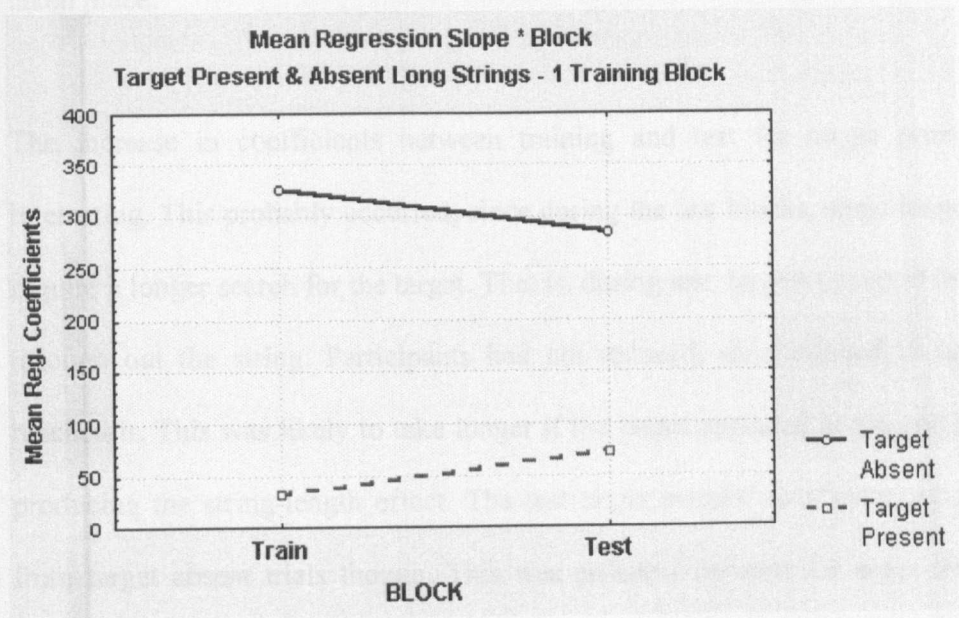
#### TARGET ABSENT/PRESENT STRINGS: STRING-LENGTH EFFECT DURING TRAINING TRIALS

As before if task redundant information becomes ignored over practice, then we should see a reduction in the string length effect over practice for trials when no target was present. On the other hand, string length should immediately have little effect when the target was present, as Ps will stop scanning as soon as it was found in the second position, no matter how long the string.

#### 1-TRAINING BLOCK CONDITION

The regression slopes calculated for each subject by block for the 1-training block condition were subjected to a 2-Way ANOVA within-subjects design 2(Target – Absent, Present. )\* 2(Block – 1,2). As indicated in figure 5.9 a main effect of Target type was revealed  $F(1,11) = 47.86$ .  $p < 0.001$ , but no main effect of Block ( $p > 0.5$ ). There was

however, a significant Target Type \* Block interaction,  $F(1,11)= 7.60$ ,  $p= 0.019$ . Post Hoc (Tukey) tests revealed that there was a reliable difference between target absent and present coefficients at both training and test as expected.



*Figure 5.9: Mean regression coefficients over Training and Test blocks for the 1- training block condition: Target Present and Absent, long strings, Experiment 3.*

Follow-up One-Way ANOVA for each Target Type finds that for target absent trials there was no reliable change in coefficients between training and test. However, for target present trials such a reliable change was found, coefficients increasing between training and test,  $F(1,11)= 32.28$ ,  $p< 0.001$ , indicating an increase in the string-length effect during the test block.

This suggests that, after one practice block the string length effect was significantly greater for Absent than Present trials. This was as expected, since after such little training,

participants should still be searching the whole string when a target was not there. However, they should stop the search immediately on finding a target at position two. For target absent trials there was no change between blocks further indicating no IR had taken place.

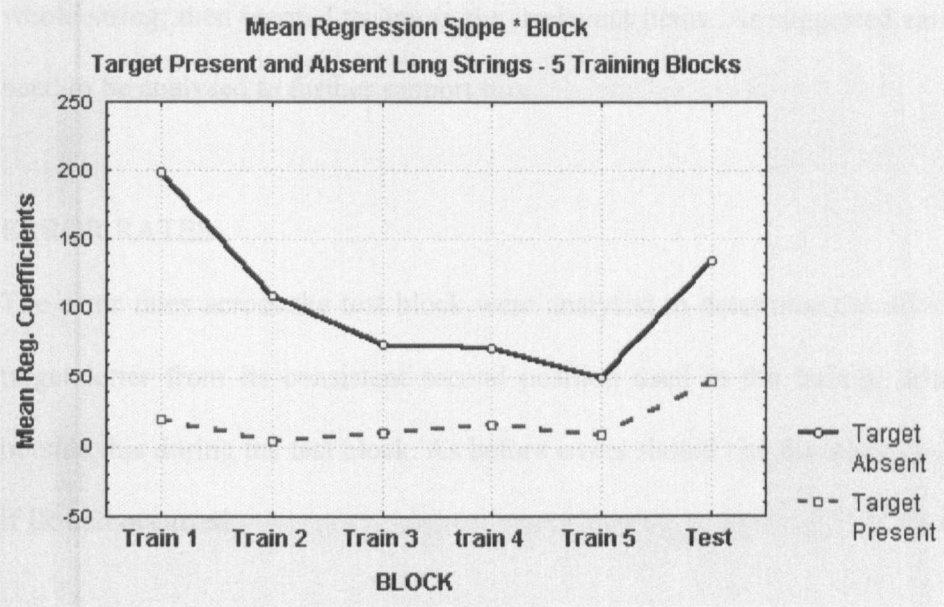
The increase in coefficients between training and test for *target present* strings was interesting. This probably occurred, since during the test blocks, some longer strings would require a longer search for the target. That is, during test, targets occurred in novel positions through out the string. Participants had not reduced, so continued to search until they reached it. This was likely to take longer if the target appeared at the end of a long string, producing the string-length effect. The test target present coefficients were still different from target absent trials though. This was probably because for some longer strings the target still appeared randomly at earlier positions, lessening the string-length effect. Overall, the 1-training block condition suggests again that IR requires practice to develop. After one training block participants were still thoroughly searching the whole string. This should be further confirmed by the later error analysis.

## 5-TRAINING BLOCK CONDITION

As can be seen from Figure 5.10, the regression coefficients started out higher for target absent than target present trials as expected, and the slope seems to reduce over trials for the target absent strings. When the coefficients for the 5-training block, target present and absent trials were analysed a 2-Way ANOVA Target(Present, Absent – within subjects) \* Block(1 to 5 – within subjects) revealed a significant main effect of Target  $F(1,11) = 25.75$ ,



$p < 0.001$ , and a main effect of Block  $F(4,44) = 11.12$ ,  $p < 0.0001$  and a Target \* Block interaction  $F(4,44) = 11.13$ ,  $p = 0.0001$ .



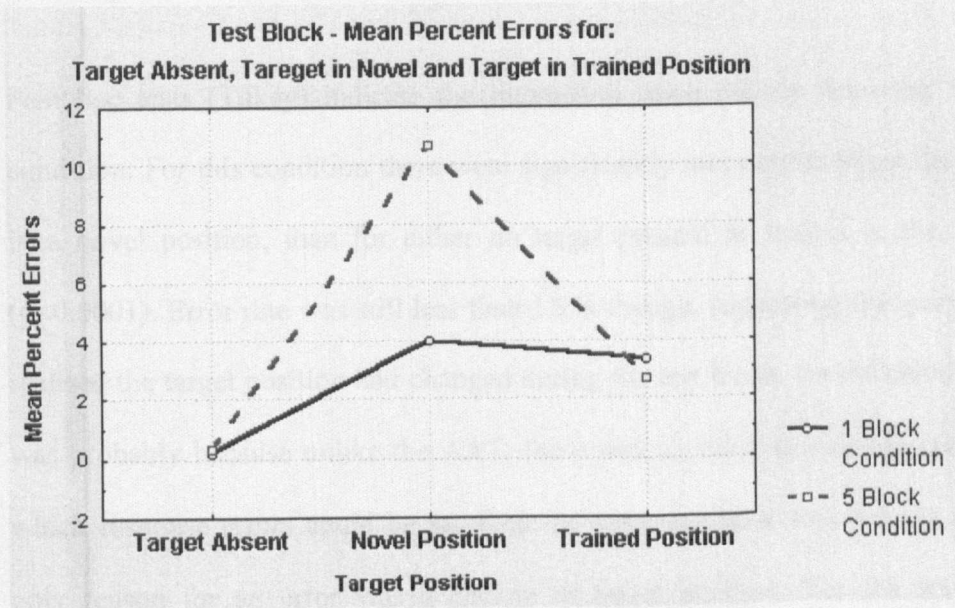
*Figure 5.10: Mean regression coefficients over Training and Test blocks for 5-training block conditions for Target Present and Target Absent strings, Experiment 3.*

A follow-up analysis comparing first to last training blocks only: 2-Way ANOVA Target(Present, Absent – within subjects) \* Block(First, Last – within subjects), found a significant main effect of Target  $F(1,11) = 27.86$ ,  $p < 0.001$ , and a main effect of Block  $F(1,11) = 14.33$ ,  $p < 0.01$  and the Target \* Block interaction was preserved  $F(1,11) = 13.74$ ,  $p = 0.01$ . Post hoc tests (Tukey) indicated this interaction was the result of a significant decrease in slope ( $p = 0.001$ ) across first to last training blocks when the target was absent. However, there is no similar reliable change in slope when the target was present.

The 5-training block data were consistent with the view that the string-length effect reduced systematically over blocks during training conditions only where the target was absent. This suggests that when there was no target, participants possibly began by scanning the whole string, then learned to ignore the irrelevant items. As suggested earlier, error scores need to be analysed to further support this.

ERROR RATES

The error rates across the test block were analysed to determine the effect of moving the target letter from its consistent second position used in the training trials to a position outside this during the test block. As before errors should rise for targets in a novel position if IR had occurred.



*Figure 5.11: Mean percent error rate for the final test blocks where target could appear in the trained or a novel position; target absent string errors added for comparison.*

Experiment 3.



As can be seen in Figure 5.11, there seemed to be greater percentage errors in the test block following five training blocks when the target occurred in a novel position compared to when either there was no target or a target in the trained position. Following one training block, overall more errors seem to have been made when targets were present than when they were not. However, in contrast to the 5-training block condition, there appears little differentiation caused by target position.

A 2(Condition – 1-training block & 5-training blocks, between subjects) \* 3(Position – No Target, Target Novel Position & Target Trained Position) ANOVA revealed significant main effects of Condition  $F(1,22) = 4.52, p = 0.045$ , and Position  $F(2,44) = 23.07, p < 0.0001$ , and a significant Condition \* Position interaction  $F(2,44) = 7.29, p < 0.01$ .

Post hoc tests (Tukey) indicate the interaction came mainly from the 5-training block condition: For this condition there were significantly more errors when the target appeared in a novel position, than for either no target present or targets in the trained position ( $p < 0.0001$ ). Error rate was still less than 12 % though, suggesting that participants quickly realised the target position had changed during the test block. As indicated previously, this was probably because unlike the AAT, there was no other calculation task in the TST to which response errors could be ascribed. In other words, it was quickly realised that the only reason for an error was a change in target position. For the one training block condition, there was no reliable difference in mean percent errors across the three different types of target occurrence. In addition to the regression analysis, this again supports the

view that following longer training, participants did learn to ignore the irrelevant string items and not just process them more quickly.

#### INTRODUCTION OF ERRORS IN THE POST-TRIPLET STRING: RESPONSE TIMES (RT).

For the final test block, following longer practice participants seemed again to make more response errors, when for target present strings, the target occurred in the untrained position within the string. As indicated in the two previous thesis experiments, feedback followed each trial, so to find the source of increased test errors, participants should resume searching the whole string. Feedback was given immediately after each trial for two reasons. First, it allowed findings to be compared with earlier published studies (Haider & Frensch, 1996) and secondly in real-world tasks feedback often occurs in this way. Admittedly, feedback in naturalistic settings can also occur some time after the completion of a task, and this will be interesting to investigate in later studies.

The change in target position should cause a rise in the regression coefficients: that is, string length should once again predict RT. As reported above, there was no reliable change in coefficients between train and test for the one block condition (Fig. 5.9), Figure 5.10 though displays a difference between these blocks following five blocks of trials.

The change in slope shown between the last training and test blocks for target absent strings in figure 5.10, for the 5-training block condition, was subjected to a 2(Target – Present & Absent, within subjects) \* 2(Block – Last Training & Test, within subjects) ANOVA. There was a significant main effect of Target  $F(1,11)= 35.31, p< 0.001$  and Block  $F(1,11)=$

40.79,  $p < 0.001$ , and a Target \* Block Type interaction  $F(1,11) = 4.90$ ,  $p = 0.049$ . Post Hoc tests (Tukey), revealed that the interaction came mainly from a reliable increase in slope for target absent trials across the two blocks ( $p < 0.001$ ), while no similar significant difference arises for target present trials.

Also, the test indicated no significant differences between absent and present trials at the last training block, but a significant difference between the two at test. Together, the increase in slope for target present and difference from target absent at test seems to suggest that scanning of the irrelevant items did re-emerge in the test block.

## RESIDUAL PROCESSING OF STRINGS

The above analysis indicates that the results of the previous experiment were replicated with the longer strings. IR seems to have progressed for the longer training condition as before. The results also seem more convincing for the longer strings and further investigation of the above analysis with post hoc tests seem to partly support this. For both the 5-training block 2-Way ANOVA's Target(Present, Absent – within subjects) \* Block(1 to 5 – within subjects) and also Target(Present, Absent – within subjects) \* Block(First, Last – within subjects)), Tukey tests found no significant difference in mean coefficient between the target present and target absent trials during the last training block.

This ties-in with the difference between training and test explained previously: there was no difference between string types for block five, but the difference re-emerged for the test block. Planned comparisons however, found that the significant difference across string types for the first two training blocks (collapsed –  $F(1,11) = 30.91$   $p < 0.0001$ ) was

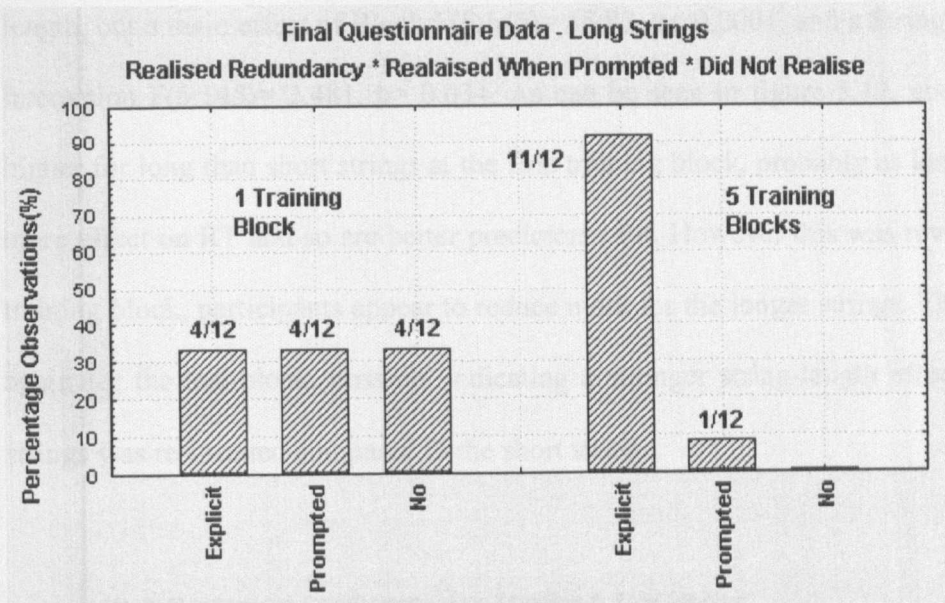
maintained at the last training block five ( $F(1,11)= 6.34, p= 0.03$ ). Suggesting, that with the extra power of the planned comparisons, a difference can be found. Overall though, this difference does not appear as robust as for the previous study.

#### END OF SESSION QUESTIONNAIRES

As before, the experimental session ended with a questionnaire aimed at determining how explicitly participants had learned the underlying regularity in the strings. Participant reports are somewhat at odds with the regression and error data this time. While this data suggests IR had not occurred in the one training block condition, 8/12 participants reported for this condition that they realised the targets appeared at a regular position, at least at some level (See Figure 5.12). Though half of these only realised the rule when prompted. For the five training block condition 90% (11/12) report that they explicitly realised the consistency of target position. The one remaining participant realised it when prompted..

Overall, it seems that increasing the length of the strings causes more participants to realise the target regularity. A number also realise it after only one block, but considering the regression data, it would appear that the confidence to use this knowledge only occurs after further practice.

Also of note is, that after five blocks in this experiment, most of the participants indicate knowledge of the regularity without prompting. Further indicating the knowledge of the rule has become explicit and consolidated.



*Figure 5.12: Participant responses to final questionnaire for Experiment 3: Those that explicitly realised post-triplet redundancy, those that realised it when prompted and that did not know of the rule.*

*(Presented as a percentage to allow comparison with Experiment 2).*

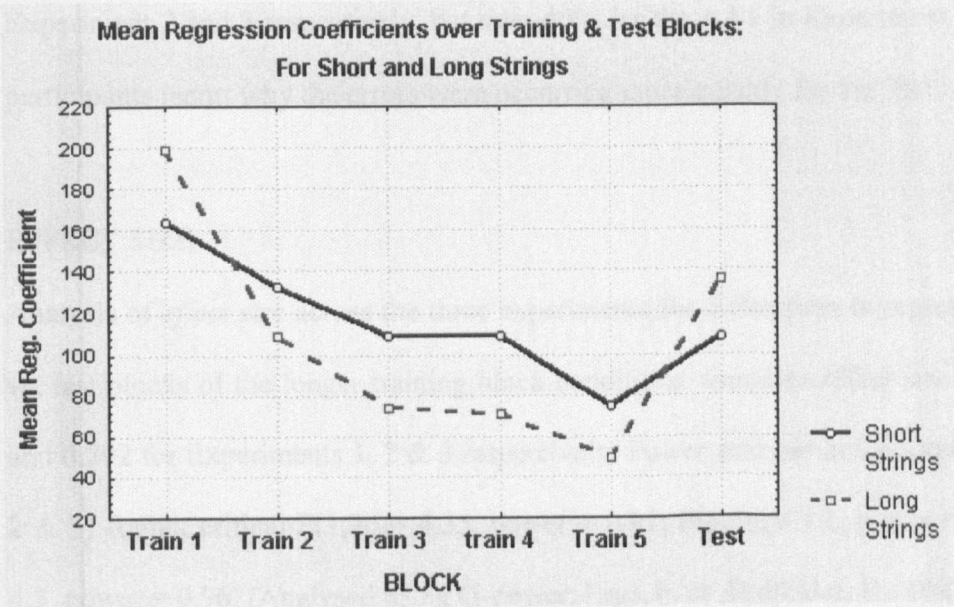
## COMPARISON ACROSS THE THREE THESIS EXPERIMENTS

A comparison across the three studies run so far would be useful to determine how the different tasks promote IR. Thesis study 2 & 3 were run almost ‘back to back’ using the same room equipment and program, only the strings were changed for longer ones. Therefore, it may still be worthwhile to compare regression slopes across the two studies.

*Statistical processing continued on the next page*

Focusing on the 5-training block conditions, a 2-Way ANOVA Strings (Short, long – between subjects) \* Block(1 to 6 – within subjects) revealed no main effect of Strings

length, but a main effect of Block  $F(5,145)= 15.83, p< 0.0001$ , and a Strings length \* Block interaction  $F(5,145)= 2.481, p= 0.034$ . As can be seen in figure 5.13, coefficients appear higher for long than short strings at the first training block, probably as longer strings have more effect on RT and so are better predictors of it. However this was reversed by the last training block, participants appear to reduce more for the longer strings. The slopes reverse again for the test block, possibly indicating a stronger string-length effect for the longer strings was re-instated compared to the short strings.



*Figure 5.13: Mean regression coefficients for the 5-training block conditions for Experiment 2 & 3: A comparison of Long and Short strings.*

Further examination of the slopes for all three thesis experiments suggests that the least residual processing occurred for the AAT, though there was some evidence that it was also reduced for the long strings experiment compared to the short strings. This indicated that

possibly IR was greater for the AAT than the TST, though lengthening the strings appeared to reduce the difference between the two different tasks in this respect.

One further interesting point was that regression slopes always seem initially higher for the shorter practice conditions, this was also reflected in the RT's. Maybe people were just more careful when they knew they had less to do, as participants were aware of approximately how long the experimental session would take. Again comparing across thesis studies, there were fewer errors in the TST than the AAT, 9% and 11% for Experiment 2 and 3 respectively, but over 40% for the AAT in Experiment 1. So it appears participants learnt why the errors were occurring more quickly for the TST.

## EFFECT SIZE

Analysis of *effect size* across the three experiments for differences in regression slope, first vs. last blocks of the longer training block conditions, found an effect size of 0.741; 0.643 and 0.792 for Experiments 1, 2 & 3 respectively. Power analysis in the same order (study 1, 2 & 3) found; critical  $F(1,20) = 4.35$ , power = 0.91;  $F(1,36) = 4.1$ , power = 0.97;  $F(1,22) = 4.3$ , power = 0.96. (Analysed using G-power, Faul, F. & Erdfelder, E., 1992). All the effect sizes were large by the normal convention (Howell, 1997, p.339).

Finally, the self-report data found that 58% realised the regularity in Experiment 1, while less than 40% did so in Experiment 2, though over 90% realised it in Experiment 3. Certainly this indicates that more people were finding and explicitly reporting the regularity of target position in the third study.

### 5.3.4 DISCUSSION

#### *Generalising the IR effect*

This experiment replicated the previous results using the TST for Experiment 2 in this chapter. The criteria set by Haider and Frensch (1996) to demonstrate IR were again met, so it can be concluded that IR does seem to occur for target search paradigms such as used here. This supports the view that IR occurs across both different tasks and manipulations of those tasks. An important consideration, if the reduction of processed information is to be accepted as a general part of skill acquisition. Before the present study there was little evidence to suggest IR occurred in other tasks. Therefore the thesis represents an important step forward in the investigation of the information reduction hypothesis put forward by Haider and Frensch (1996).

The strings in Experiment 3 once again did not repeat within blocks of the TST. This further strengthens the argument that an instance account such as Logan's (1988) cannot explain the improvement in performance (unless the concept of an instance is relaxed to include some abstraction across instances, such as the best method of task accomplishment – see literature review Chap. 2, Sec. 2.2.4). This was significant, since mnemonics appeared to be frequently formed during the AAT, but it now seems, were not necessary for IR to progress. Thus, learning repeated instances was not required for IR to develop. Since IR has also been shown to persist across different stimulus-sets (Haider & Frensch, 1996), this now rules out an 'instance' explanation at any point during either the learning or deployment of the reduction strategy. Consequently, this is further evidence that contemporary theories of skilled performance will need to incorporate IR.



### *Task difficulty*

The aim in increasing string lengths was to increase task difficulty and determine if this then increases IR compared to the shorter strings. Taken as a whole, the results do suggest slightly more IR for the longer strings of Experiment 3 compared to the previously used short strings (for the longer practice conditions). Participant self-report supports this, explicit statement of the regularity increased for Experiment 3. One possible explanation, for this increase in reduction across experiments, is that the addition of extra string items caused some change in response criteria setting. This is a similar argument to that put forward by Haider and Frensch (1999a) and involves random walk theory.

### *Random walk theory and IR*

Random walk theory has been invoked by Haider and Frensch (1999a) to explain the effect of task instruction: instructions to work quickly promoted IR, whereas instructions that emphasized accuracy suppressed the development of IR. In the rational view proposed by Haider and Frensch (1999a) the setting of more liberal response criteria as set-out in random walk models is most likely to be *accompanied* by a reduction in the processing of task irrelevant features. Thus, to some extent random walk theory and in particular criterion-setting provides an explanation of IR under the influence of task instructions. This may be important for the results of the thesis studies since Ratcliff (1981, 1985) has reported that not just task instructions, but also task difficulty, may moderate the setting of response criteria. Accepting this, the trend in IR caused by increasing the TST string length may possibly be explained in terms of criterion-setting.

Before examining this, though, the way in which random walk theory may be applied to IR overall should be explored in greater detail and the term ‘accompanied’ unpacked. That is, in the typical IR task what information is ‘getting in’ to the random walk process and how exactly does this apply to the data presented here.

### *Haider and Frensch’s (1999a) use of Random Walk Theory*

To explain their data and the effects of task instruction Haider and Frensch (1999a) have drawn on a substantial body of research investigating decision making, especially in simple two-choice decision tasks. Firstly, it is known in the response-time literature that individuals can adjust both their speed and accuracy in responding when instructed to do so. The usual explanation for this is that different response criteria are used following the influence of task instruction (Pachella & Pew, 1968; Wickelgren, 1977). Haider and Frensch (1999a) suggest that the setting of such response criteria is often described within a random walk process, particularly citing Ratcliff (1981) and his diffusion random walk model.

Ratcliff’s (1978) random walk model (RWM) is similar to a number of others (such as Laming’s (1968) earlier description of such processes) and is perhaps one of the most successful in capturing a variety of experimental data (Cohen & Nosofsky, 2003). To sketch out a description of a RWM, imagine a simple two-choice reaction time experiment in which a probe is presented and the task is to decide if it is one of a memory-set or not. In the model itself (fig 5.14.) there would be a ‘starting point’ from which information begins to be accumulated. Above this would be a ‘match boundary’ (or match response criterion) and below the start point would be a ‘non-match boundary’. Leaving out a number of

specifics that will become important later in the discussion, the idea is that information is gradually accumulated (either constantly – as in a diffusion process, or in discrete steps). As this happens a ‘count’ is kept of the number of feature matches and non-matches and ‘drift’ occurs from the starting point towards one or other of the two boundary criteria. If enough information is accrued to drive the drift to one of the boundaries a response is made. For example, a ‘match-response’ may be made if the probe is one of the memory-set and this information drives the drift to the match criterion. In these models, particularly Ratcliff’s (1978, 1981), both the criteria boundaries and the starting point are variables that can be set by the individual.

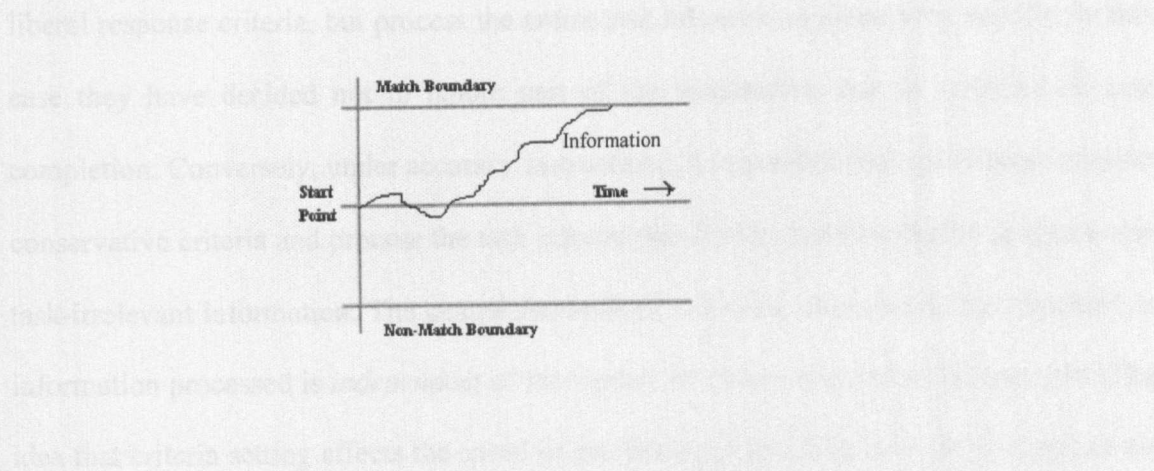


Figure 5.14: Information accumulation within a Random Walk Model

When using RWM’s to explain RT data under speeded instructions (instructions to work as quickly as possible), the speed-accuracy trade off can be modelled by setting more liberal response boundaries. That is, bringing the boundaries closer together, so that less information is sought before a response is initiated. This not only makes for a faster response, but also can lead to an increase in error rate; this follows the empirical data, validating the model somewhat.

It would appear that such a conceptualisation may fit with IR under the influence of task instruction; as both talk about processing less information under speed stress. This is a view Haider and Frensch (1999a) also find attractive in their article. They do at the out-set highlight a caveat that the setting of response criteria may be independent of a decision over *which* information is processed. However, they also go on to suggest two views at this point; the independence view and the rational view.

*The independence* view outlines a situation where under speed stress participants adopt liberal response criteria, but process the entire task information albeit very quickly. In this case they have decided not to ignore part of the information that is irrelevant to task completion. Conversely, under accuracy instructions, it is possible that participants may set conservative criteria and process the task information slowly, but now decide to ignore any task-irrelevant information. The central rationale of this view, then, is that the ‘amount’ of information processed is *independent* of the ‘speed’ at which information is processed. (The idea that criteria setting effects the speed of processing is probably incorrect though, as the drift rate towards a boundary would be unaffected only the *amount* of information changes before a response, not the speed of accumulation).

*The rational view*, which Haider & Frensch (1999a) favour takes a different view. In this they suggest that the setting of liberal response criteria would be *accompanied* by limiting task processing to task-relevant components. So under speeded instructions participants may choose to ignore task-irrelevant information on the basis of little evidence, and subsequently quicken performance. In the conclusion of their paper they seem to buy into

the language of random walk theory further. They suggest, that under speed instructions, participants in their study moved their response criteria closer to the starting point and so made a response on relatively little information. In contrast, again in line with their results, they suggest that under accuracy instructions the criteria boundaries are moved further away from the starting point so that relatively more information is extracted from the task stimulus.

So, they argue their results suggest that setting response criteria not only affects the speed of processing (more likely to be the speed in reaching a response), but also *which* information is accumulated to generate a response. This they indicate explains the data in the language of random walk. They then re-describe the findings in the language of IR: when under speed pressure participants try to reduce the information they process. That is, once they recognise the irrelevant information, they try to ignore it to speed performance. In a final statement in the paper it is reported that under a rational analysis the findings support the view that the setting of response criteria and the reduction in processing irrelevant information should not be viewed as independent (Haider & Frensch, 1999a).

There are a number of fine points to this argument, but it does seem to be implying that information reduction can be explained, at one level, as the setting of response criteria within a random walk model. As indicated, there seems to be quite some evidence that under speed pressure people respond faster, but with more errors. This is well modelled by a RWM when the response boundaries are brought closer to the starting point. Thus, RWM's appear to account for both the response time data in the literature and the IR results when task instructions are manipulated to stress either speed or accuracy. Haider & Frensch

(1999a) seem to suggest from this that when liberal response criteria are set, there is an accompanying systematic reduction in processing of task-irrelevant stimulus components. The big question is whether IR can be explained overall by invoking a random walk process?

The problem in deciding this comes from two related directions. Firstly, determining what is the object and what are the features and secondly, there is some lack of specificity about the perceptual input to RWM's. Taking the last issue first, a number of the papers on decision making which use random walk models say little about 'what is getting in' during the information accumulation stage. Often terms are used such as 'evidences' (Rouder, 2000) or 'features' (Ratcliff, 1978) or even a 'stream of information' (Laming, 1968).

It is not entirely clear what information is entering the random walk (RW) process. In a similar way, while Ratcliff (1978, 1981, 1985) talks of each item in a string comparison task consisting of a number of features, Lamberts, Brockdorff and Heit (2003) treat the stimulus holistically. In this latter case, it is the number of *whole* exemplar comparisons from memory that increment the RW process and not the accumulation of matching and non-matching features suggested by Ratcliff (1978). This highlights the second problem: How can it be determined across various tasks what is an object and what are features – a current problem in attention research and covered in the literature review. This is perhaps a major weakness of RWM's; what gets into the 'comparator' may vary from task to task so much that the models have little generality.

*Applying Random Walk Modeling to IR paradigms*

Accepting this rather loosely specified input within RWM's, an IR process may fit within the RW concept. For instance, imagine the whole strings of the Alphabet Arithmetic Task (AAT) or the Target Search Task (TST) are taken as one stimulus.

Example of AAT and TST	
<i>Alphabet Arithmetic Task (AAT)</i>	E [4] J K L M (when the 4 letters are skipped does the whole string follow the alphabet).
<i>Target Search Task (TST)</i>	Memory set - C, O, V Search set - F V N H G W P (is one of the memory set present in the search set)

The stimulus string then requires one response decision of pressing button A or button B to verify the stimulus. If each string item is regarded as a 'feature', then speeded instructions may result in fewer of these features becoming processed as response criteria become more liberal. At this point it sounds very much as if IR can be re-described as a RW process, in which the items of information are the individual string items. However, further consideration of IR in terms of RW processes finds the approach to be wanting.

The first problem is the way information would be reduced in a random walk. Ratcliff (1978) does indicate that certain features may be arranged so they are processed first. This would be expected in say a word-matching task, where the letters would be read from left to right automatically. Narrowed response boundaries would mean these preferential early features would be processed while the later ones would not. Thus, there may be some

preferential processing of less information in a random walk. However, this is not IR, the reduction strategy laid out by Haider and Frensch (1996). They describe the isolation of task-relevant from task-irrelevant information and then the strategic decision to ignore the irrelevant information. This is the systematic reduction in the information processed and seems to be beyond the simpler or automatic processes described within a RWM. It appears that the reduction of information processed under liberal criteria in a RWM is not the same information reduction described in the IR hypothesis, at least in the contemporary description of the models.

A bigger problem is that Ratcliff (1978) is quite clear that RW processes model a single stage decision process with response times around 1.5 seconds maximum, not decision processes that require multiple stages (Ratcliff and Rouder, 2000). To set-up the argument both the AAT and TST were described as single decision tasks as they are two-alternate forced choice paradigms. To allow this, the strings were taken as one holistic stimulus, which is obviously not the case (Cohen & Nosofsky, 2003). Both tasks require multiple decisions within them. The AAT requires both an arithmetic calculation and the separate decision over whether each successive letter in the string follows the alphabet. Even the simpler TST requires each string letter to be compared separately to the memory set to determine if a match is present. So, in these two tasks we do not have the single stage decision process that Ratcliff and Rouder (2000) are describing. Subsequently, the decision over which response to make often took longer than 1.5 seconds, particularly in the AAT.

The TST used in this chapter is quite similar to the Sternberg paradigm described by Strayer & Kramer (1994) and Ratcliff (1978) in their modeling. Both papers are quite clear



over what is happening in the RW process: for *each* comparison between memory set and probe, a parallel feature comparison takes place. The information accrued from this comparison increments a counter or causes drift towards one or other of the response criteria (absorption boundaries). So it seems clear that the RW process takes place within each string-item to memory-set comparison and not across the string items. Thus, the final decision which overall response to make is a *subsequent* process (Lamberts, Brockdorff & Heit, 2003).

It would seem then that if the typical RWM is invoked in a description of the AAT or TST, its scope would be to describe how decisions are made on comparisons between the items in memory and individual string items. To summarise, it has been found that under speed pressure task-irrelevant items may be more likely to be ignored. Additionally, it is likely that fewer individual *item features* may be processed if such speed stress causes the setting of more liberal response criteria. However, the reduction in information processed in the random walk is not the reduction of information in IR; most likely they are two separate processes.

#### *Random walk theory and TST data*

Random walk decision processes, if accepted as a valid model, undoubtedly could occur in the TST. Each string item could be compared to the three items in the memory set in the way described by both Ratcliff (1978) and Strayer and Kramer (1994). How this process may be affected by the change in string length cannot be directly determined from reviewing the data in the studies in this chapter. However, when targets were within the trained position, errors were close to just 2% for both the shorter and longer TST strings, so

it would seem that these items were compared quite thoroughly and were not subject to any changes in a random walk evidence accumulation process.

It was only when targets appeared in untrained positions that the error rate increased (comparing figure 5.4 with 5.10 also shows slightly more mean errors for longer strings). Thus, while there were possibly a number of different practice effects occurring, the error data for targets in the untrained position during the test block demonstrate that IR, as described in this thesis, was occurring. Most likely, IR made the largest contribution to the change in response times over practice and the length manipulations.

### *IR and criteria setting*

This is not to say that both processes – RW and IR, are not influenced by ‘criteria setting’. As Haider and Frensch (1999a) suggest and to use their terms, rationally, the change in one will *accompany* the other. However, most likely they are independent processes. For example, in a task with no irrelevant information, speed pressure may cause fewer *item* features to be processed, but all the stimulus items to be attended. Conversely, in tasks that allow IR, fewer items may be processed, but each item processed fully if it is not possible to separate each into individual features. The relative contribution of IR and RW within a task may thus be open to a number of task specific factors. In the case of both the AAT and TST, given sufficient speed stress a combination of both could occur.

It appears that IR is not really captured within contemporary RWM. However, the language of RW is useful, particularly the concept of criteria setting. Treisman and Williams (1984) report a theory of criteria setting that is separate from random walk modelling; theirs is a

two-process model of criteria setting. The first process sets the response criteria in the general region using past experience and referents. This process may be affected by task instructions and by other task parameters, such as difficulty. The second process involves fine-tuning the criteria based on recent experience or history. Such a concept may help to understand the conditions that may moderate IR. Considering that criteria setting may influence both the random walk and IR processes, task dimensions shown to affect the former, such as task difficulty and stimulus probability could be worth investigating within an IR paradigm

### *Task load and criteria setting*

Within the TST, it seems that response criteria (for target present and target absent responses) may be set more liberally when perceptual load was increased (Lavi and De Fockert, 2003; Lavi *et al.* 2004). Explained in another way, as the strings got longer and so harder to process, IR became more worthwhile and so more likely to be adopted. A similar view is proposed by Newton & Roberts (2000) to explain why a cancellation strategy was only used when pencil and paper were unavailable to simplify their compass task. When the task could not be written down this made the problem more difficult, as all the steps had to be retained in memory, this encouraged simplifying strategies to be developed and used. Increasing string length seems to increase IR, and the most likely explanation for this is that as the task became more difficult, strategies were used more consistently to reduce the load.

### *Task attributes*

There is something of a paradox here though; in that IR arises more completely in the AAT than the TST, the two only becoming similar when the TST strings were made much longer

(accepting that a direct comparison within one experiment would be better to decide this than the comparison used across experiments). What perhaps was so surprising with the AAT was that people were prepared to ignore the short post-triplet segment to such a degree. This segment contained a maximum of 4 letters and all they had to do was determine whether they followed the alphabet. Recognising the alphabet is an over-learned skill, so this was very easy, yet people were still prepared to ignore this part of the string. Thus, it seems there may be attributes of the AAT that make it particularly prone to IR. As already suggested this may be because the task was relatively (compared to the TST) difficult and boring, or there may be other factors that have an impact

Increasing the length of the strings was chosen as it directly affects the cost/benefit of reducing. To not reduce for the longer strings meant that up to 11 more letters needed to be scanned. Under Lavi and De Fockert's (2003) criteria, this made the task more difficult as perceptual load was increased. Lavi (1995, 2000) makes a distinction between such perceptual load and cognitive load: proposing cognitive load can also affect attention as it increases. Given that cognitive load can be conceptualised as Working Memory (WM) load (Lavi, 1995, 2000), this is an interesting distinction, since WM load was not really increased for thesis study 3. The difference in WM load may be one reason for the apparent difference between the two tasks.

Strengthening this argument, the AAT has been used to directly investigate WM and age effects (Lincourt et al, 1998; Brigman & Cherry, 2002). Typically, to increase WM load the digit was ascended. For example, J[2]M or L[4]Q, as before the digit indicates letters to skip. The task, as usual, was to verify if the string followed the alphabet when the requisite

number of letters was skipped. The latter string required more letters to be skipped mentally, making the task more demanding on WM limitations.

Brigman and Cherry (2002) investigated this further, presenting only the triplet part of the string in their task (as IR was of no interest in this case). They found a correlation between age, WM capacity and performance at the task, and proposed WM mediates the retrieval of stored instances and this mediation declines with age. What this does show is that the AAT loads WM. It is possible that this aspect of task difficulty makes IR very likely in the Haider and Frensch (1996) version of the task. The next experiment reported in chapter 6, was intended to explore this and understand more fully the type of ‘task attributes’ that may moderate IR.

## CONCLUSION

The present study provides evidence that IR generalises to a different task; the target search task. This further supports the view that such reduction strategies may be a general part of skilled performance. Of particular interest was the role of instance learning in the development of IR. The design of the TST, used in both experiments for this study, tends to rule out the necessity of this type of learning. These two points together strengthen the argument that IR needs to be integrated into contemporary accounts of skill acquisition.

The following chapter investigates whether certain task attributes, such as string repetition and task difficulty, whilst not necessary for IR might still moderate the effect. This will not only further determine the conditions under which a reduction strategy is most likely, but also the contribution of task-intrinsic factors within this.

## **CHAPTER 6: STUDY 3**

### **IR and Moderating Task Attributes**

*This chapter examines the role that different intrinsic task attributes may play in moderating the information reduction (IR) strategy. Additionally, levels of anxiety and problem solving ability were measured, to determine if individual differences along these dimensions, distinguish those more or less likely to adopt the reduction strategy.*

## 6.1 INTRODUCTION

Information reduction (IR) appears to be a strategy learned without explicit cues about task irrelevant items. The strategy also seems to be employed consciously by individuals. Further, IR does not seem to be stimulus specific, which is problematic for some instance accounts of skill acquisition (Haider & Frensch, 1996, 1999a). Reduction also seems to be moderated by extrinsic factors, such as task instruction (Haider & Frensch, 1999a). This short summary suggests knowledge about the mechanism of IR is increasing. However, much of what is known about IR has been gleaned from one task, the alphabet arithmetic task (AAT). Because of this, investigations of task attributes that may moderate the effect have been limited. It is possible though that there are a number of both task intrinsic and extrinsic properties that may affect the use of the strategy.

In psychological experimentation, small task manipulations can sometimes completely change the result. For instance, Croner and Albright (1994, 1997) found that when noise dots in a Random Dot Kinematogram display were in a different colour or polarity to target dots that moved cohesively, thresholds for motion detection fell. That is, in a noisy display of dots moving in random directions, a sub-set of dots all moving in the same direction are more easily spotted if they were different in some dimension. However using only a slightly different technique, with noise and targets at different luminance, Edwards and Badcock (1994) found this segmentation provided no advantage for motion detection. There is, therefore a contradiction between these two studies.

The critical difference between the studies was that in the first case ‘all’ the targets were one colour/polarity and all the noise was the other colour/polarity. However, in the later

experiment, a proportion of the noise shared the colour/polarity of the targets. In this case even though the 'set' size was much reduced by the segmentation, there was no advantage for motion detection thresholds. The reason proposed for this (Snowden & Edmunds, 1999) was that when all the targets were a colour/polarity not shared by the noise dots, a high level attentional strategy could form.

This strategy allowed focus on just one or two target dots, of a particular colour/polarity, to track their motion. However, such a strategy could not work effectively if some of the noise dots shared colour/polarity with the target dots. The disparity therefore arose between the two original experiments since in one case a cognitive strategy could be used, while it was much less successful in the other. The point is, one small change in the target/noise attributes gave very disparate results. This case is pertinent, as it describes the formation of an attentional strategy that serves to reduce the amount of information processed.

In a similar way, as discussed earlier, Lavi (Lavi, 1995, 2000; Lavi, Hirst, De Fockert & Viding, 2004) reports that evidence supporting both 'early' and 'late' explanations of selective attention can be resolved by understanding the differing perceptual and cognitive loads arising in the experiments. For example, experiments with high perceptual load mean little resource is left over to process unattended stimuli, accounting for the early selection findings. Conversely, low perceptual load would mean processing 'spills' over to unattended items that then have to be 'filtered out' at a later processing stage, explaining the late selection results. This again illustrates not only how disparate results are often caused by small shifts in task parameters, in this instance perceptual load, but also how such discrepancies lead to a better understanding of underlying mechanisms.



## TASK ATTRIBUTES

Considering that IR is proposed as part of a theory of skill acquisition (Haider & Frensch, 1996), and such theories are likely to have an applied role, it is perhaps even more important to determine how IR operates across varying tasks and task manipulations. Realising the conditions which mediate the development of IR is necessary to understanding its role in everyday skilled accomplishment. Further investigation of task attributes that may moderate IR therefore seems useful line of enquiry.

Haider and Frensch (1999a) have already shown task instruction as a moderator, IR increasing under speeded instructions. The previous two experiments in this thesis have also shown IR to occur in an additional task, further generalising the effect. The two experiments in chapter 5, together also suggest some effect of task difficulty: more pronounced reduction occurring for increased string lengths. This raised something of a paradox though. The redundant items in the TST are scanned more when processing load was low: with short strings. However, considering the AAT in comparison, the post-triplet verification task was not only very easy and of the same length as the TST short strings, but was really an ‘over-learned’ alphabetic recognition task. So it is very surprising that individuals chose not to check this part of the string throughout AAT.

This highlights the possibility that there maybe some attributes of the AAT that increase the extent of IR. The most parsimonious explanation is that the increased cognitive load of the triplet calculation discourages post-triplet scanning compared to the TST. Therefore, the effect of task difficulty would seem to warrant further investigation.

### *String learning*

Of particular interest, in addition, is the possible contribution that instance learning (Logan, 1988, 2002) might make to information reduction. While study 2 in this thesis indicated that such learning was not *necessary* for IR, instance learning within the AAT may provide a reason why this task did seem to result in robust reduction. Robust, even when the irrelevant items were very simple to process. Evidence consistent with the possibility that memory-based retrieval was occurring was found in the first experiment of the thesis. Participants reported the use of mnemonics to avoid the triplet calculation. Inevitably this led to the realisation for some that strings were repeating, and may even have promoted reduced processing of the post-triplet segment.

Haider and Frensch (2002) also report that some participants appeared to have learned components of the strings in their studies. This could occur because constraints inherent in the version of the AAT they used meant that at most only ten correct and ten incorrect unique triplets were presented (Haider & Frensch 1996, 1999a, 1999b, thesis study 1). Thus, instance learning could have occurred in both Haider and Frensch's experiments, as well as our own. In terms of the theoretical debate surrounding the role of rules and examples in skill acquisition (Anderson, Fincham and Douglass, 1997), there is at least some evidence then that specific examples, or instances, may be committed to memory in the AAT. Subsequently, these instances are most likely retrieved during the performance of the task. If individuals realise strings are repeating and recognise them as they are presented, this may inevitably increase confidence in the rule (post-triplet items may be ignored), leading to direct retrieval of the correct response from memory.

Taking this a little further, there is converging evidence that strategy selection is moderated by a feeling of knowing (Schunn *et al.*, 1997, Miner & Reder, 1994, Reder 1987). The strategy choice process may be an integral part of answering problems (Reder, 1987). Schunn *et al* (1997) suggests that in a straightforward choice between retrieving an answer from memory or computing the answer, the chosen strategy depends on a familiarity based feeling-of-knowing process. That is, when faced with a problem, people initially decide whether to compute a solution or retrieve it from memory depending on how familiar that problem feels. Further, Reder and Ritter (1992) found that this familiarity judgement could be influenced by the form of the problem, such that a choice to retrieve was still made for novel problems if the form was similar to learned instances. For example if the original learned problem was  $1+2*3+2$  and a novel problem was  $1*2+3*2$ , participants still made the decision to retrieve rather than compute, even though attempts at retrieval lead to incorrect answers

If this is applied to the AAT, such a feeling of familiarity may increase IR. If participants realised that the number of distinct triplets is relatively small and that they repeat, they may attempt to memorise them. As a result, they may gain confidence in the triplet as a reliable source for string verification. The triplet could then become a flag for the retrieval of the correct answer. Such a successful strategy allows participants to abandon both triplet computation and the scanning of the rest of the string. So in a similar way as suggested by Schunn *et al* (1997), the switch to an IR strategy may be dependent on a feeling of familiarity with the problems.

Of course, the IR effect has been shown to transfer to novel stimulus sets (Haider & Frensch, 1996), in which case simple retrieval will not be successful. However, the form of the problem stayed the same across sets. The feeling of problem familiarity may have

encouraged participants continued confidence in reducing to the triplet as the verification source. It would then aid the generation of the correct response, as possibly indicated by Reder and Ritter (1992). While such an attribute as instance retrieval may not be necessary, as indicated with the TST, and the transfer study (Haider & Frensch, 1996), realising the strings repeat may increase confidence in abandoning a level of residual processing for the post-triplet string. This may then explain why following practice, there was little residual processing of irrelevant items within the AAT used for study 1. Even though the alphabet recognition sub-task was over-learned and relatively easy, confidence in the rule was high enough to ignore this segment.

It is possible then, that reducing the number of instances in a task may moderate the overall degree of IR. This should provide an interesting experimental variable, especially since instance learning is central to a number of theories concerning skilled performance (Logan, 1988; Gobet, 1996).

### *Task difficulty*

The TST studies also suggest that task difficulty may have such a similar mitigating effect. As the strings got longer and the task load (or at least perceptual load, Lavi & De Fockert, 2003) increased, so to some extent did IR. Additionally, Ratcliff (1985) proposed that task difficulty may change the response criteria or boundaries within his random walk model of response time. So as a task gets more difficult, upper and lower response boundaries are set closer together by the individual and a response is made after less information is gathered. Such changes in criterion setting may also influence IR, if this is the case, as the difficulty of a task increases so should IR. Therefore, it may be interesting to investigate the relative contribution of task difficulty on the overall extent of IR, particularly the effect of cognitive load.

### *Perceptually separate information sources*

There has been some discussion concerning cueing and information selection (Ball & Sekuler, 1981; Chun and Jiang, 1998, 1999; Olson and Chun, 2001): under cued conditions relevant information is more efficiently located and processed. It would seem logical then, that some cue to relevant information would encourage the progression of IR. In the Haider and Frensch (1996, 1999a, 1999b) experiments there were no instructional cues as to the relevant information in the AAT. However the triplet did form a focus for attention and was a separable region both perceptually and as a sub-task. Experiments have shown that triplet saliency is not a key factor in IR (Green & Wright, 2003; Lincourt, 1998). However, these experiments still used strings that formed separate task ‘chunks’. So perceptual or cognitive separation of the string may arguably still have had some effect.

Finally, IR occurred in the TST, so cueing does not necessarily underlie reduction, but may promote it. Understanding the conditions under which IR is most likely to occur, and the extent to which it may occur, is just as important as providing evidence of the overall effect. This is particularly central considering that IR may prove an important factor in everyday tasks. Understanding the task criteria under which IR is likely to be a predominant part of skilled performance may be important both in teaching complex skills and improving efficiency in skilled task performance.

## **6.2 EXPERIMENT 4: INFORMATION REDUCTION AND TASK ATTRIBUTES**

In summary, three factors that may moderate the overall degree of IR within a task have been identified. These were;

- (i) the number of unique task instances,
- (ii) overall task difficulty and
- (iii) perceptual separability of ir/relevant information.

The following experiment in this chapter sought to examine the relative contribution that these three factors might make to the overall degree of information reduction. Since there was room for further reduction with the TST short strings, these were again used in the same form and then modified to produce the experimental conditions laid out below. The TST was initially developed so that it differed from the AAT on a number of possibly key dimensions, the next experiment could be seen as sequentially re-introducing attributes of the AAT, such as separability, difficulty and repetition, to ascertain their contribution to the effect. The hypothesis was that IR (as indicated by test block error data and declining string-length effect over practice) would be greatest in Condition 3 and least in the Standard condition. The exact contribution of each attribute remained to be determined.

### **6.2.1 METHOD**

#### **DESIGN**

A three factor mixed design was used - Error Type (Trained-Position vs. Post-Trained-Position – within subjects) \* String Length (five levels, 3,4,5,6 & 7 items – within subjects) \* Condition ( Standard, 1, 2 or 3 – between subjects). The dependent measures were Mean error rate per block and RT. The experiment was designed specifically to determine the effect of different task attributes in the three experimental conditions compared to the standard strings from Experiment 2.

## PARTICIPANTS

Participants were 48 staff and students of the Open University, Milton Keynes, with an age range of 23-60 yrs. and a mean age of 41. Participants were randomly ascribed to one of the four conditions. Each person received £10 for participating in the experiment.

## STIMULI AND PROCEDURE

The stimuli and procedure were very similar to Experiment 2 chapter 5. Participants were tested separately in a quiet room. Each person sat in front of a PC monitor and keyboard, viewing strings of letters approximately 0.4\*0.4 cms in size and 0.4cms apart, e.g. the Standard Condition took the form of a display of varying numbers of letters such as

J F K N E T L

As before the experiment was generated and run in E-Prime, (for technical specifications see Schneider, W., Eschman, A., & Zuccolotto, A., 2002a, or Schneider, et.al, 2002b), running in Windows 98SE at 1024\*768 resolution.

Participants responded using a standard PC keyboard by pressing keys marked as;

- 1 for memory-item present and any adjacent letters were after the target in the alphabet.
- 2 for memory-item present and any adjacent letters were before it in the alphabet or.
- 3 if the memory-items were absent.

This was reversed for half the participants. The Standard Condition and Condition 1 differed from this; in only having keys marked for 'Target present' or 'Target absent'. Again the keys were reversed for half the participants. Each string was presented after a 500ms fixation cross and was displayed until a response was made. The strings, which were presented in blocks of 100, were made up as follows.

**i. Standard Condition.** Participants verified 5 training blocks and one test block of strings. In each training block there were 50 trials where the memory target item (F, M or T) was present at the second position from left, and 50 trials where the target item was absent. The string length varied from 3 to 7 items long. Ten examples of each length were generated for both present and absent conditions (repeated to create the five training blocks). The strings themselves were randomly created using a Visual Basic program as before, with F, M or T in the second position for half the strings and absent for the remaining 50 strings.

The test block was generated as the training blocks, except that the target now appeared 25 times in the usual location and 25 times at a location further right than the second position. The position of the targets 'outside' was counterbalanced as far as possible across the 25 strings. As before participants were not informed of the regularity or the change between training and test blocks

**ii. Condition 1.** This differed from the Standard condition in that brackets were used to segment the strings. As before the memory item was always located at the second from left position, but now was contained in a bracketed set of three items which preceded the rest of the set of 1-4 items e.g.,

[J F K] N E T L.

If any participant asked what the brackets were for (n=1) they were told it was to make the experiment more similar to an earlier one.

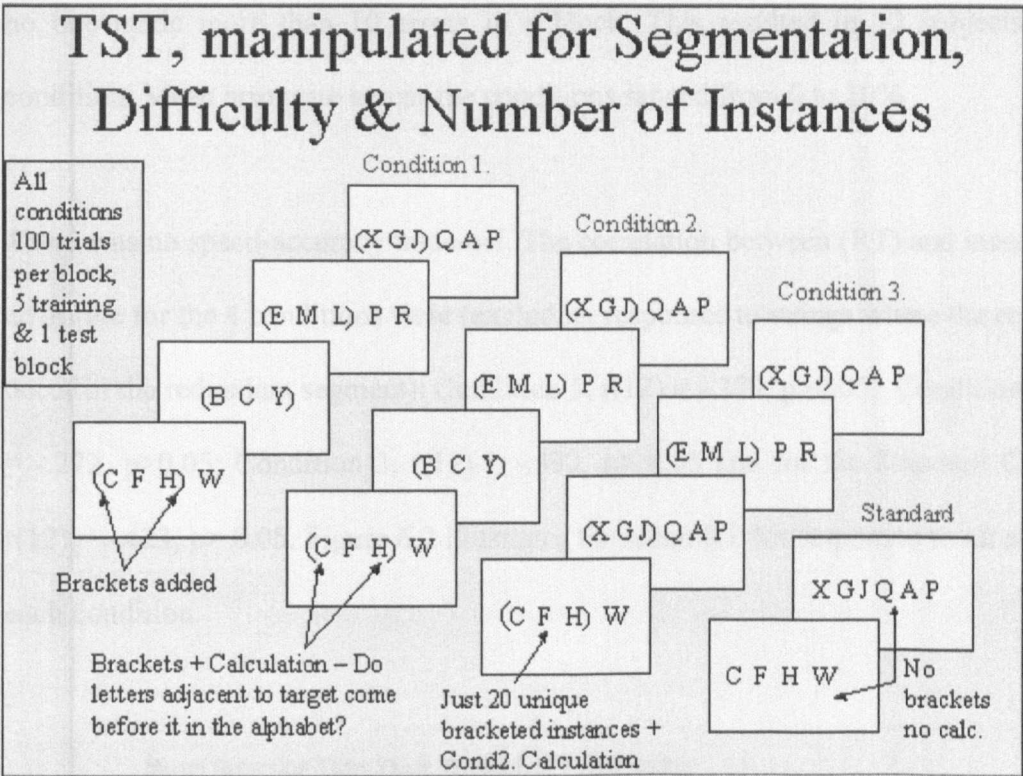


**iii. Condition 2.** This used strings from Condition 1, which had also been created so that 25 of the target present trials had letters adjacent to the target item that came before it in the alphabet. In the remaining 25 ‘present’ trials, letters adjacent to the target came after it in the alphabet. The 50 absent trials were as Condition 1. The test block followed this pattern, though sometimes the target appeared outside the brackets (for half the target present trials); the adjacent letter calculation was still required.

**iv. Condition 3.** This used strings similar to Condition 2, except now just 10 unique target-present and 10 absent strings, each containing 3 items (just the bracketed set) were selected from Condition 2. To each, between zero to four letters were added to generate 50 instances of target-absent strings and 50 of target-present strings.

For all conditions participants read and signed an informed consent form (See Appendix 1), then the session proper started. Each participant attempted ten practice strings repeatedly until accuracy was at least nine correct, after which the experimenter left the room. The participants then completed five training and one test block. For the Standard and Condition 1, the task was a simple two-alternate forced choice, deciding if a target was present or not. For Condition 2 & 3 the task was more difficult. If the target was absent, there was as before just one key to press, however if the target was present there were a choice of two further responses. If the target was present adjacent letters had to be taken into account. If these letters came before the target in the alphabet one button was pressed, but if they came after in the alphabet an alternate marked button was pressed. The description sounds more difficult than in actual performance and all participants found the task understandable by the end of practice. At the end of the six blocks, the experimenter returned to the room and asked some questions to determine if an awareness of the rule had occurred. At the end of this participants were debriefed,

and a note was made of any comments they chose to make. The different conditions are further illustrated in figure 6.1 below.



*Figure 6.1: Example of the strings presented for each of the conditions for Experiment 4.*

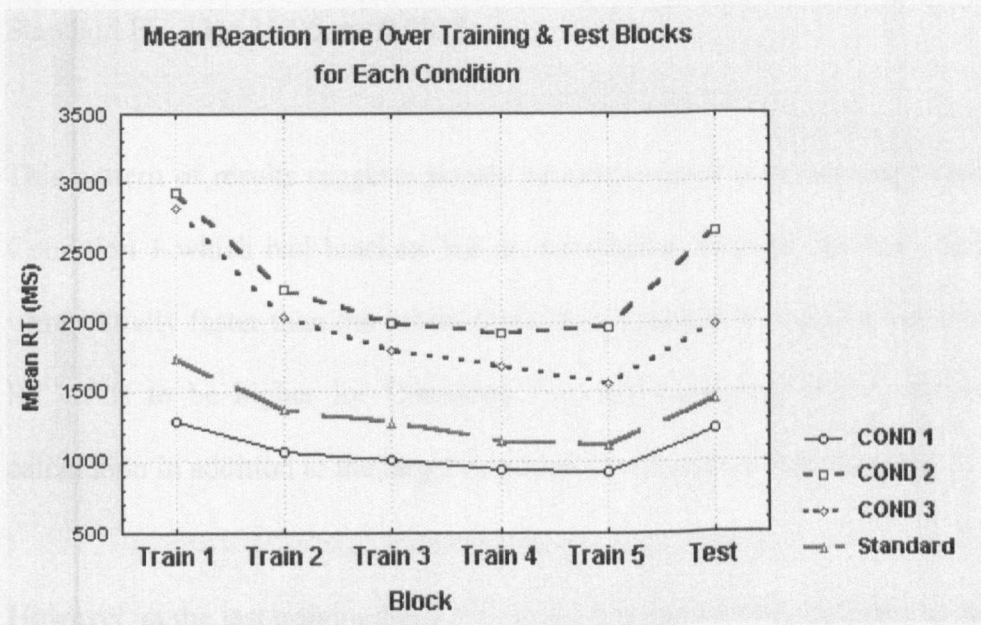
**6.2.2 RESULTS**

**OVERALL ANALYSIS**

As previously, to enable the effect of string length on RT to be compared directly, for all conditions the best fitting linear regression slopes were computed separately for each subject and each trial block. The regression coefficients obtained were then subjected to ANOVA. The analysis followed the same form as the previous studies in the thesis, although this time for the four conditions.

Mean error rates were computed for each subject in each practice block, subjects were excluded from the analysis if their error rate was higher than 10% throughout the training blocks (N = 0). Surprisingly, even for the conditions with the added calculation, no one made more than 10 errors in a block. This resulted in 12 subjects in each condition. Mean error rate across the conditions ranged from 0 to 10%

There was no speed-accuracy trade-off. The correlation between (RT) and mean percent error rate for the 4 conditions were (excluding responses to strings where the error could occur in the redundant segment): Condition 1,  $r(12) = -.279$ ,  $p > 0.05$ ; Condition 2,  $r(12) = -.272$ ,  $p > 0.05$ ; Condition 3,  $r(12) = -.392$ ,  $p > 0.05$  and for the Standard Condition,  $r(12) = -.423$ ,  $p > 0.05$ . Figure 6.2 illustrates the mean RT for responses to all strings for each condition.



*Figure 6.2: Mean Response Time (RT) for string verification over blocks for the four conditions, Experiment 4*

As can be seen in Figure 6.2, RT reduces over training blocks for the four conditions, 2-Way ANOVA Block (1-6, within subjects) \* Condition(1-4, between subjects) revealed a main effect of Block  $F(5,220) = 54.66$ ,  $p < 0.0001$  and a main effect of Condition  $F(3,44) = 6.09$ ,  $p = 0.0015$ . There was a Block \* Condition interaction  $F(15,220) = 4.029$ ,  $p < 0.0001$ .

Post hoc tests (Tukey) revealed a significant reduction in RT between the first and last training blocks ( $p < 0.05$ ) for all conditions except Condition 1. and a reliable increase in RT between the last training and test blocks ( $p < 0.05$ ) for Condition 2 & 3 but not Conditions 1 & Standard. It was predicted that there would be a group difference at the end of training. Planned comparisons for the last training block revealed that compared to the Standard Condition there was no reliable difference in RT for Condition 1 or Condition 3. However, Condition 2 was found to have statistically higher RT than Standard  $F(1,44) = 11.06$ ,  $p = 0.0018$ .

This pattern of results suggests people became quicker over training blocks except for Condition 1 which had brackets but no calculation. Though, for this condition people were initially faster than the others ( $p < 0.05$ ), so have less room for improvement. That RT tends to be higher for Condition 2 is not surprising; this condition required a calculation in addition to the target detection for Condition 1 & Standard.

However, at the last training block Condition 3 is not reliably different to Standard. This indicates that the same calculation was having less effect on RT for this case. The most parsimonious explanation is that people were able to learn the reduced set of instances in Condition 3, so could retrieve the answer and avoid the time consuming calculation. In effect, they then became closer to the Standard condition in response times.



INFORMATION REDUCTION WITHIN THE CONDITIONS: REGRESSION SLOPES, TARGET ABSENT STRINGS

The change in mean regression coefficients over training and test blocks is shown in figure 6.3 The slopes for target absent trials appears to decline over training blocks for all conditions and rise again for the test block as expected.

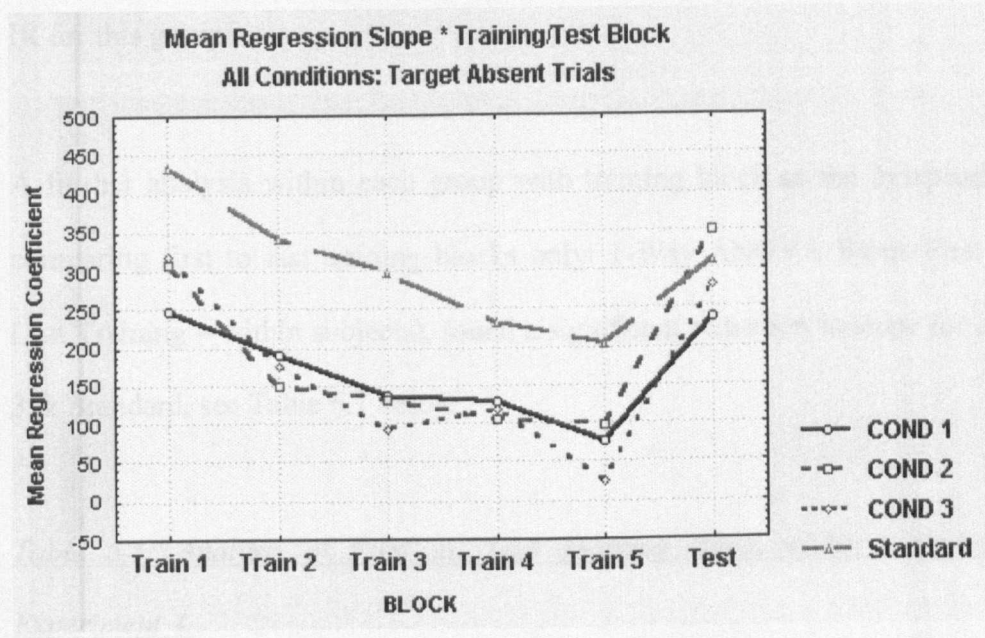


Figure 6.3: Mean regression slopes over Training and Test blocks for all conditions: Target absent trials, Experiment 4.

The data was investigated further with a 2-Way ANOVA Condition(1-3 & Standard – between subjects) \* Block(1 to 6 – within subjects). This revealed a significant main

effect of Block  $F(5,220)=35.13$ ,  $p<0.0001$ , but no main effect of Condition. There was a Block \* Condition interaction  $F(15,220)=1.88$ ,  $p=0.026$

Follow-up separate One-Way ANOVA for the first training block found no reliable difference between groups. A similar analysis of the last training block however, did find a main effect of Condition  $F(3,44)=3.01$ ,  $p=0.04$ . Post hoc tests (Tukey) indicated the only significant difference was between the Standard condition and Condition 3 (reduced number of instances)  $p=0.027$ . Comparing conditions again for the test block finds no reliable difference between conditions. This suggests that at the last training block string-length was having least effect on Condition 3, indirectly signifying greatest IR for this group.

A further analysis within each group with training block as the independent variable comparing first to last training blocks only: 1-Way ANOVA Block(First Training & Last Training – within subjects), found a significant reduction in slope for conditions 1-3 & Standard, see Table 6.1 below.

*Table 6.1: Analysis of First vs. Last Training Block within each condition, for Experiment 4.*

Condition	Main effect of Block	first / last training block
Standard	$F(1,11)=19.510$ ,	$p < 0.01$
Group 1	$F(1,11)=35.217$ ,	$p<0.0001$
Group 2	$F(1,11)=23.655$ ,	$p<0.001$
Group 3	$F(1,11)=15.142$ ,	$p<0.01$

The regression slopes analysis suggests that the string-length effect declined over training for each condition, which was indicative of IR. If the change was due to IR, then the trend should reverse for the test block, during which reduction must lead to errors.

The slopes in Figure 6.3 seems to support this, as they all rise during the test phase. To determine if the change was reliable, within each condition the last training block was compared to the test block: One-Way ANOVA's Block(Last Train, Test – within subjects) were again run. The results in Table 6.2 below indicate that the rise in regression coefficients was significant for all groups.

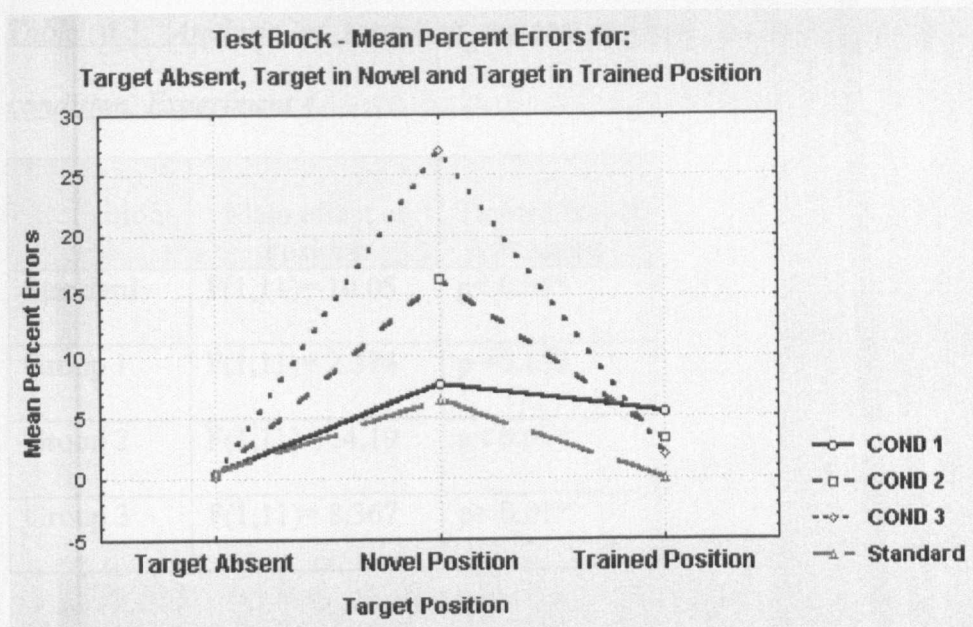
*Table 6.2: Analysis of Last Training vs. Test Block within each condition, Experiment 4*

Condition	Main effect of Block	last training/test block
Standard	F(1,11)= 9.688	p < 0.01
Group 1	F(1,11)= 70.344	p<0.00001
Group 2	F(1,11)= 31.443	p<0.001
Group 3	F(1,11)= 73.662	p<0.00001

ERROR RATES

The error rates across the test block were again analysed as previous experiments to determine the effect of moving the target letter from its constant second position in the training trials to a position outside this in the test block. As before errors should rise for targets in the novel position if IR occurred. It was hypothesised that the experimental manipulations would moderate IR, if this is so the error rates should vary across conditions.





*Figure 6.4: Mean percent error rate for the final test blocks where target could appear in the trained or a novel position; target absent string errors added for comparison, Experiment 4.*

There did appear to be a difference in percentage errors between conditions when the target appears in a novel position. 2-Way ANOVA Condition(1-3 & Standard – between subjects) \* Position(Trained, Novel & Absent – within subjects) found a significant main effect of Condition,  $F(3,44)= 3.325$ ,  $p = 0.028$  and Position  $F(2,88)= 28.306$ ,  $p < 0.00001$ . Also there was a significant Condition \* Position interaction  $F(6,88)= 4.155$ ,  $p < 0.01$

Within the conditions, follow-up ANOVAS comparing percentage errors for targets in Trained versus Novel positions found significantly more errors when the target occurred in the novel position for all conditions except Condition 1 (see table 6.3 below)



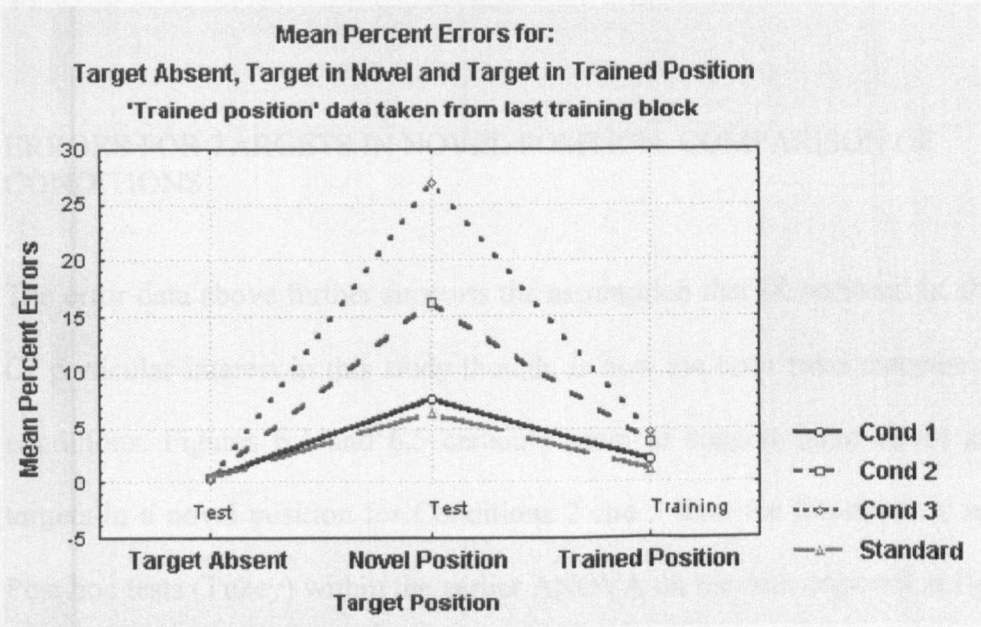
Table 6.3: Analysis of target at trained position vs. novel position within each condition, Experiment 4.

Condition	Main effect of Position	Trained/Novel Position
Standard	F(1,11)= 10.05	p< 0.01*
Group 1	F(1,11)= 2.374	p =0.152
Group 2	F(1,11)= 14.19	p< 0.01*
Group 3	F(1,11)= 8.367	p= 0.01*

One possible ad hoc reason why error rate failed to increase by position for Condition 1 is that the task was as easy as the Standard conditions and the string was split into two segments. Upon receiving error feedback there may have been alternate switching of attention back and forth from relevant to irrelevant segments. This would explain why there were nearly as many targets in the usual position missed as targets in the novel position. That is, people were not sure where to look so started to miss both target types. To test this assumption a little further the error rates of the last *training* block for targets present (always in the expected position) were compared to targets present in the novel position during the test phase.

An identical analysis to that above, but substituting ‘*training* target present’ data for *test* phase target present in the trained position, found a very similar result See figure 6.5 and table 6.4 below). A similar 2-Way ANOVA found again a main effect of Condition, F(3,44)= 4.50, p <0.01and Position F(2,88)= 27.72, p< 0.00001. Also again, there was a significant Condition \* Position interaction F(6,88)= 3.377, p< 0.01

At this stage, all conditions when analysed separately show more errors when the target was in the novel position compared to the trained position. This tends to support the proposal that for Condition 1, the confusion caused by errors in the test phase meant attention was alternately varied between relevant and irrelevant segments. This caused more targets in the trained position to be missed than would be expected from the previous training block.



*Figure 6.5: Mean percent error for target in the trained or a novel position; target absent string errors added for comparison, for Experiment 4. (Error data for targets in the trained position taken from the last training block).*

*Table 6.4: Analysis of target at trained position vs. novel position within each condition, For Experiment 4: (trained position data taken from last training block).*

Condition	Main effect of Position	Trained/Novel Position
Standard	F(1,11)= 6.941	p = 0.023
Group 1	F(1,11)= 12.66	p <0.01
Group 2	F(1,11)= 12.956	p< 0.01
Group 3	F(1,11)= 6.566	p= 0.026

ERRORS FOR TARGETS IN NOVEL POSITION, COMPARISON OF CONDITIONS

The error data above further supports the assumption that IR occurred in all conditions. Of particular interest in this study though, is how the error rates compare between the conditions. Figures 6.4 and 6.5 certainly seem to suggest more errors are made for targets in a novel position for Conditions 2 and 3 than for Condition 1 and Standard. Post hoc tests (Tukey) within the earlier ANOVA on the data depicted in figure 7.4 find that for targets in novel positions only Condition 3 (reduced number of instances) has reliably more errors than the Standard condition (p< 0.05)

Follow-up ANOVA's ( 2(Condition – between subjects) \* 3(Position – Absent, Novel & Trained – within subjects)) where just one condition at a time was compared to the Standard strings, found that both Condition 3 and Condition 2 (increased difficulty) have higher percentage error rates than the Standard condition for targets occurring in untrained novel positions when tested post-hoc via Tukey tests (See table 6.5).



*Table 6.5 Comparison of error rates between Standard and one other condition separately – post-hoc (Tukey) comparison of errors for novel target positions, Experiment 4.*

Condition Vs. Standard	Main Effect of Condition	Main Effect of Position	Interaction	Post hoc comparison of errors for targets in novel position
Cond 1.	F(1,22)= 3.21, p= 0.087	F(2,44)= 19.01, p< 0.0001	F(2,44)= 3.34, p= 0.044	p= 0.95
Cond 2.	F(1,22)= 11.80, p= 0.002	F(2,44)= 28.71, p< 0.0001	F(2,44)= 5.36, p= 0.008	p= 0.001*
Cond 3.	F(1,22)= 5.55, p= 0.028	F(2,44)= 12.71, p< 0.0001	F(2,44)= 4.93, p= 0.01	p= 0.003*

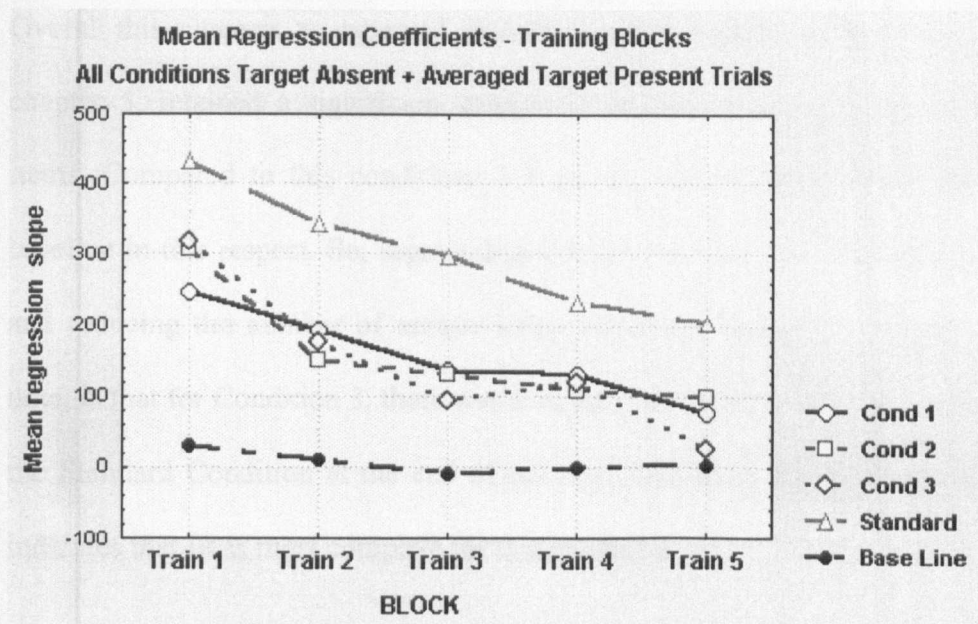
Overall it seems that when the number of instances were reduced for Condition 3, there was a significant rise in errors for test targets appearing in untrained positions. This error level was reliably different from the Standard condition. Less convincingly, there does seem some evidence that increasing task difficulty (Cond 2) also increased these errors from the Standard group. Segmenting the strings with brackets appeared to cause no reliable difference from Standard in this respect.

### RESIDUAL PROCESSING

As in the studies in the previous chapters understanding how complete IR becomes by the end of training can be achieved by comparing the mean regression slopes in the last training block for target present and target absent trials. When a target was present the search was most likely terminated at the second position from left during training. This then represents what the target *absent* slope should become if reduction was complete.

An average of the target present trials for all conditions was added to the analysis to form such a baseline comparison. The data are represented in figure 6.6 below.

comparisons confirm this pattern of results.



*Figure 6.6: Mean regression slopes for all conditions over blocks compared to a baseline formed from the average of the target present trials collapsed across groups, Experiment 4.*

determining how explicitly participants perceived the target.

Two-Way ANOVA Condition(1-3, Standard & Baseline – between subjects) \* Block(1 to 6 – within subjects) found a significant main effect of Condition  $F(4,55)= 6.11, p< 0.001$ , and Block  $F(4,220)=, p< 0.0001$ . There was also a Condition \* Block interaction  $F(16,220)= 2.24, p= 0.005$ .

Post-hoc (Tukey) tests revealed that by the last training block, the only condition that differed significantly from the baseline was the Standard condition ( $p< 0.001$ ) which retained a reliably higher regression coefficient than the target present baseline measure.

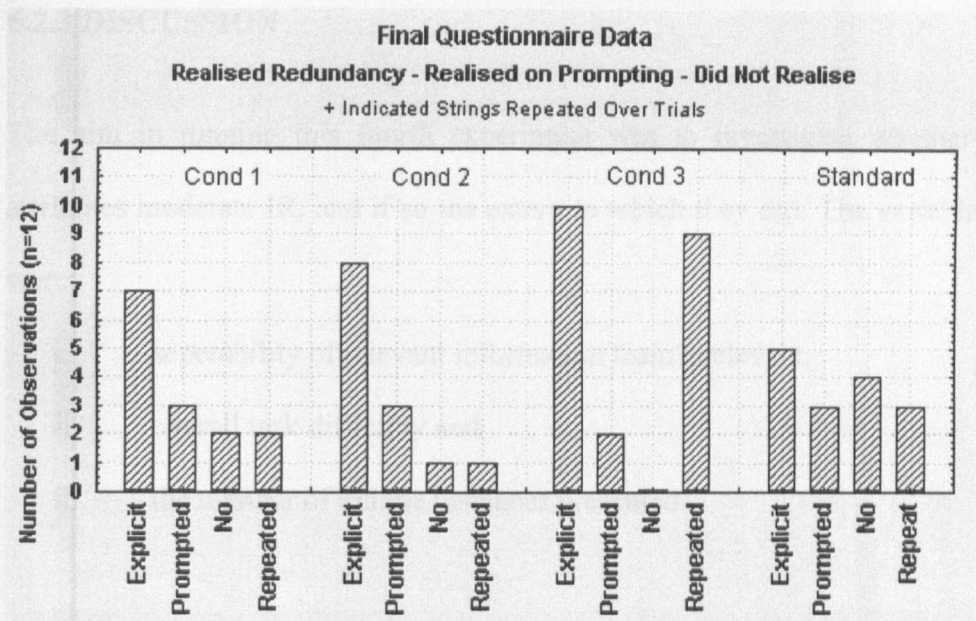
Additionally, as mentioned earlier the only condition that differed reliably from the Standard condition at the last training block was Condition 3 ( $p = 0.0011$ ). Planned comparisons confirm this pattern of results.

Overall this suggests, as expected, that the standard condition taken from Experiment 2 chapter 5, retained a significant amount of residual processing for redundant string items. Compared to this conditions 1-3 do not reliably differ significantly from the baseline in this respect. So, segmenting relevant information, increasing task difficulty and reducing the number of unique string instances all seem to enhance IR. Also of note, is that for Condition 3, there was a significant reduction in coefficient compared to the Standard Condition at the end of training. This taken together with the error data indicates that IR is most complete for this condition.

## END OF SESSION QUESTIONNAIRES

As with the previous experiments, each session ended with a questionnaire aimed at determining how explicitly participants had learned the underlying regularity in the strings. That is, those that realised during training, that targets appeared at a consistent position; relegating the string items past this position irrelevant to task completion.





*Figure 6.7: Participant responses to final questionnaire: Those that explicitly realised post-triplet redundancy, those that realised it when prompted and that did not know of the rule, for Experiment 4. (Number that indicated they thought the strings repeated over trials also added).*

The Qualitative data in figure 6.7 appears to follow the earlier analysis of the experimental conditions. The highest number of people reporting explicit knowledge of the rule was for Condition 3 (10), the next highest Condition 2 (8), then Condition 1 (7). The lowest number reporting this was the Standard Condition (5). While in conditions 1, 2 and Standard a few participants (falsely) thought the strings repeated, by far the highest number was for Condition 3, where they actually did repeat. This supports the view that participants were aware of this manipulation. As before a small number of people seemed to realise the regularity when prompted.

### 6.2.3 DISCUSSION

The aim in running this fourth experiment was to investigate whether certain task attributes moderate IR, and if so the extent to which they did. The variables considered were;

- i.       seperability of relevant information from irrelevant,
- ii.       overall task difficulty and
- iii.      the number of unique instances presented.

The relative contributions to IR of seperability and increases in task cognitive difficulty were unknown. It was anticipated from past results that participants would try to commit instances to memory where possible.

#### INTRINSIC TASK ATTRIBUTES

Whilst it is true that none of these attributes are necessary for IR to occur, it is possible that they influence the extent of IR. One of the surprising findings within the AAT was that redundant information appeared to be consistently ignored, even though it involved a relatively easy task-recognition of an alphabet sequence. Recognising the alphabet can be considered an over-learned task and there were at most four extra letters to process, so why were they ignored so completely? In answer, it is likely that there are a number of task factors that may promote IR. The question posed here is did intrinsic task attributes of the AAT make IR particularly likely.

##### *Number of unique instances*

When the number of unique instances in the target search task (TST) were made equivalent to the AAT, the results seemed clear. At the last training block the mean



regression slope approached zero, indicating that there was very little residual processing of the redundant items. Indeed, when an average target present condition was added to the analysis, there was no reliable difference between it and the target absent slope for this condition at the last training block. This suggests that the redundant items were having very little effect on response times. Indirectly, this indicates a high level of IR.

Examination of the error rates for the test phase strengthens this argument. When targets occurred outside the usual position, errors increased significantly for all the groups (though arguably not that well for Condition 2), it was apparent though that these errors were much higher for Condition 3. This suggests that participants were continuing to concentrate on the first few items even after they had received some error feedback indicating incorrect responses. Again this points to a high level of IR that was persisting through the test phase. Overall, the data for this condition not only indicates strong reduction, it looks very similar in degree to the first experiment of the thesis using the AAT. Thus, reducing the number of unique instance does appear to moderate IR.

From a theoretical perspective this was interesting. Haider and Frensch (1996, 1999a, 1999b) have set-up a description of IR that does not fit well with contemporary theories of skilled performance. Particularly, it seems theories reliant on an instance account of learning (Logan, 1988; Gobet & Simon, 1996) do not allow for IR. This is specifically evidenced by the last of Haider and Frensch's 1996 experiments, which found transfer of IR across different stimulus-sets; strict instance-based theories cannot explain this (unless the concept of an instance is revised to include some abstraction about the task). However, the current experiment suggests that instance learning may affect IR. While there was no report of mnemonics being used by participants, it does seem that the

repeating instances were noticed. At the end of the experiment, eight out of the twelve participants in Condition 3 reported that the strings were repeating.

The RT data also supports the occurrence of instance learning; the alphabetic task calculation had less effect on RT when the number of unique instances was reduced to a learnable number. An explanation for this was that the correct answers were retrieved, rather than calculated following practice. So, in one task there seems to be an effect of both IR and possibly instance learning. These somewhat contradicting views can however be resolved if IR and instance learning are seen as two separable strategies, that can occur and interact within a task.

The main thrust of Logan's (1988) instance theory is that as practice progresses, a switch from calculating task solutions to retrieving them from memory occurs. Skilled speeded performance is therefore the reflection of this switch to one-step retrieval. Such a change can be imagined for many tasks and seems entirely possible for the AAT itself. Of course for such a switch to happen, instances must repeat for learning to take place. In the current target search experiment, when all the strings are unique IR still occurred, so instance learning was not necessary. However, when the number of unique instances was few and could be learned, this was sufficient to moderate IR. It seems likely then, that not only was an attempt made to learn string instances to avoid the calculation, but so doing increases IR.

This seems straightforward, but within the strings of Condition 3, it was the bracketed relevant segment that repeated frequently across trials in a block. The irrelevant letters varied much more, so that 50 correct and 50 incorrect strings could be realised. Subsequently, it is only when attention was focused on the relevant segment that the

reduced set for this condition provided an advantage. So it would appear that the relevant part of the instance was selectively learned. Therefore, it would seem that some level of reduction happens followed by a realisation that the relevant items are repeating, which further bolsters IR.

Returning to Logan's (1988) instance theory, this suggests how it and IR can be combined. As Logan states, it is what is attended that gets into an instance. However, the exact content of an instance is poorly defined at present and as considered in the literature review could feasibly even contain some abstraction about a task (Section 2.2.4). Also poorly defined is the mechanism controlling attention (outside of a weighting for current goals). Since there is no substantial definition of how an instance is formed, perhaps then, IR could be seen as one of the factors that guides attention. In doing this IR may help determine the content of an instance. This could be an interesting area for future research, revealing one of the mechanisms that guides attention..

Taking another learning account, Gobet and Simon's (1996) template theory also relies on stored, structured information to explain expert performance, in this case chess skill. Instances are held to be organised into chunks and these connect to form templates. It is the employment of these templates that explains the superior performance of skilled chess players. In a more recent paper (Campitelli & Gobet in press), there was some suggestion that irrelevant stable stimuli may be ignored when chunks/templates are activated, but the exact mechanism that facilitates this was not specified. Again some means, such as IR, to direct attention to relevant information is required.

While it is accepted that instance learning is a root explanation for expert and even automatic performance, it does seem that the selection of the content of each exemplar through attention requires further specification. It also seems that IR is fundamentally an attentional strategy – attention is withdrawn from irrelevant items and focused on relevant areas. As such it is a candidate for one of the ways in which the content of each instance may be selected.

Integration of IR into current theories of skilled performance is still a long way from resolved; perhaps a more immediate question is how the repetition of instances in the TST served to moderate IR. Two promising answers were introduced earlier. One possibility is that the repeating instances may modify the familiarity of the strings. It has been suggested that the choice between retrieving the correct answer and calculating it, can depend on a feeling-of-knowing, a processes that is directly related to the familiarity of the stimulus (Schunn *et al.*, 1997, Miner & Reder, 1994, Reder 1987). Applying this to the current study, as the relevant string segment repeated, familiarity was likely to increase and so might the realisation that the irrelevant items that go with it never contained a target.

### *Problem familiarity*

Under these conditions, the level of IR becomes more understandable if it is accepted that string verification becomes just retrieval of the relevant segment. That is, when the strings become familiar enough a switch is made from analysing the whole string to just retrieving the answer to the bracketed items. The necessary return to scanning the whole string at the test stage was also less likely; the retrieval processes was faster and the strings were very familiar. So realising something has changed takes longer, explaining the increased error rate.

The role of string familiarity provides a good explanation for the pattern of data in Condition 3. However, it falls short of explaining the rest of the results in both this and previous studies. Firstly there is evidence of IR in all groups in the study reported here. For three of the conditions the strings can only be familiar in form; the actual string items constantly vary across trials. Therefore, string familiarity is a weak explanation for general IR. Secondly, increasing task difficulty and segmenting the relevant information also seemed to reduce the degree of residual processing compared to the Standard condition. In addition to these main points, there was a trend suggesting that manipulating task difficulty increases errors for test targets at untrained positions. Overall, this indicates that it is not just familiarity that moderates IR. Indeed as already reported, task instruction to be speeded or accurate also has a marked effect on such reduction. (Haider & Frensch, 1999a). Finally, saying that familiarity effects strategy choice does not really explain how familiarity functions in information reduction.

## CRITERIA SETTING

To explain the effect of task instructions upon IR Haider and Frensch (1999a) employed random walk models (Ratcliff, 1981, 1985). To reiterate, in the most basic form random walk models describe how information is gathered until a response choice is made. The interesting aspect of such models is the proposition that information is gathered until an upper or lower response criterion is met and the subsequent response is initiated (see previous discussion outlined in p221 in chapter 5 and figure 5.13). Critically, these two response criterion can vary both in their distance from each other and the starting point of the information search. If boundaries are set closer together less information is sought before a response. Conversely if they are set further apart more information is collected before acting. As discussed in the last chapter, criteria setting within the

random walk process would most likely be linked to a change in the processing of task-irrelevant information.

Applying the language of Ratcliff's (1981) diffusion model to the task instruction results explains how speed instructions may increase IR, while stressing accuracy reduces it. From a 'random walk' perspective, an instruction to work quickly encourages more liberal response criteria than under accuracy conditions. In other words, response criteria are positioned closer to the starting point of information accumulation and so less information was sought before either a positive or negative response was made. One caveat here though, as pointed out by Haider and Frensch (1999a, pp 132), is that setting liberal criteria may not result in IR per se, the whole stimulus may still be processed, but less thoroughly. As they suggest though, a more rational view is that an increasingly liberal criteria should make irrelevant items less likely to be processed.

While IR is not really explained by random walk processes (an argument clarified in the previous chapter p221), the differential setting of response criteria (separate criterion for positive and negative responses) does seem to particularly explain the IR speed-accuracy data of Haider and Frensch's (1999a) earlier study. If it is accepted that varying response boundaries so they are closer together increases IR, while increasing the distance between them does the converse, then this implies that criteria setting influences both random walk processes and IR in the same direction.

To support the generality of the random walk diffusion model, Ratcliff (1981, 1985) has shown fits to more than just speed/accuracy data. Within the response-time literature the model appears to fit results for changes in task difficulty, stimulus probability,

brightness discrimination, masking experiments and lexical decision (Ratcliff, 1985; Ratcliff & Rouder, 2000; Ratcliff, 2002; Ratcliff, Gomez, & McKoon, 2004). It is possible that the setting of a positive and negative response criterion could be a fruitful way to understand the conditions that inhibit or facilitate ignoring task-irrelevant information – such as string repetition.

In the language of such models, familiarity or at least repeating instances may also serve to relax information criteria, so that the upper and lower criterion for a response are set closer together or closer to the start point (zero information). If this is the case, it would explain why there seemed to be less residual processing compared to the Standard Condition strings, since less of the redundant information will be sought. Closely set boundary criteria would also predict more errors for novel placed items at test. It seems plausible that instances that appear to be consistent and repeating would result in a less strict response criterion.

One thing that was implicit in Ratcliff (1981) description was that the boundaries were set as a conscious decision. This view supported by the Haider and Frensch (1999a) experiment, which found IR varied within subjects when task instructions were changed. When instruction were switched between ‘accuracy’ and ‘speeded’ task directions, IR was moderated, suggesting some conscious control of the deployment of the strategy. In a similar way for the current experiment, participants did report that the strings were repeating and overall a number reported the consistency of target position. So most likely, they were consciously using this information to set response criteria. However, whether Ratcliff’s (1981) model fits data where familiarity is manipulated is still an empirical question. Theoretically at least it appears possible.

### *Criteria setting and increased task difficulty*

The concept of criteria setting when task difficulty is manipulated is less satisfactory as the direction of change found here may not match what is expected from random walk (RW) modelling. To re-cap, the current data indicate the following pattern. Less residual processing was found when task difficulty was increased compared to the standard strings. However, the drop was no greater than when only brackets were added. There did seem to be a trend for more novel position test targets to be missed when standard and difficulty conditions are compared. However, this was not as convincing as when the number of unique instances was reduced.

Thus, the results for task difficulty are equivocal, but do hint at increased reduction compared to the easier strings. Further supporting this direction of change, the findings reported in the previous chapter, suggests that increasing the string length encourages IR. Arguably, this was also the result of an increase in overall task difficulty. So it would seem plausible that increasing task difficulty tends to further reduce processing of task-irrelevant items.

Examining Ratcliff's work (Ratcliff, 1981, 1985; Ratcliff & Hacker, 1981) reveals that task difficulty not only affects response times in his studies, but is accurately modelled by changes in boundary position and search start points. Specifically, when the string matching decision becomes more difficult, the *different* boundary is moved further away from the start, slowing response times. What is apparent from this description is that as the response decision becomes more difficult, response boundary/s are moved further from the starting position of an information search. This implies more information was sought in these conditions. Abstracted to the IR paradigm, this would likewise indicate more information, particularly the redundant items, would be



processed as the task required more effort. Clearly, this would result in a different pattern of data than found here. Far from increasing the processing of redundant items, increased task difficulty tends to cause the converse when compared to the standard easier strings.

Of course Ratcliff's results are for a different type of task, one that has no redundant elements, but this does raise questions over how applicable the link between criteria setting in RW and IR really is. One could propose that within the IR paradigm, increasing difficulty causes boundaries to be brought together to ease processing load. Taken to an extreme, very difficult tasks would necessitate narrowing the distance between response boundaries to alleviate unachievable task demands

To some extent, making the task either difficult or more speeded is similar in this respect; IR increases to reduce task load. IR was found to be particularly high when extra speed pressure in the form of reduce stimulus display times was used (Haider & Frensch, 1999a, expt 2) It seems a rational suggestion that increased difficulty will increase a tendency toward IR, but this does highlight a problem with the diffusion model and criteria setting in general. While such models can fit data and explain a range of findings, there is little prediction which manipulations will move the boundaries; the starting point or the direction of change. To some extent the idea of criteria setting becomes circular and post-hoc

A case in point is the *brackets only* group, there is little residual processing by the end of training, but not the expected rise in errors for test targets at novel positions. If liberal response criteria had been set, explaining the former it does not explain the latter finding. As an aside, the provision of brackets could also explain the drop in residual

processing for the *difficulty* group, but not the error pattern at test. Of course, it may be that the criteria are changed very quickly at test for the relatively easy ‘brackets only’ group. Again though, this suggests a need for further constraints on the properties that set such boundaries.

In defence of criteria setting as an underlying mechanism of IR, as indicated in the previous chapter, there is a range of literature that suggests such a phenomenon underlies performance (Pachella, 1968; Rabbitt, 1989; Sperling & Doshier, 1986). Treisman and Williams (1984) set out the beginnings of a theory of criteria setting involving both long and short-term memory processes. In particular they suggest short-term traces are used to track expected stimulus probabilities and optimise response criteria setting. Additionally Strayer and Kramer (1994a, 1994b) express the view that setting more liberal response criteria may be one characteristic of advanced skilled performance. IR would seem to fit with these ideas.

## SUMMARY

The current study further replicates IR using the TST, finding redundant items are ignored to some extent in all the conditions. The *Standard* condition, which uses strings similar to the earlier study in the thesis, found a similar level of performance once again. Compared to these strings, manipulating segmentation of the relevant information, increasing task difficulty and reducing the number of unique strings all seem to effect reduction.

The most convincing difference occurred when the number of unique instances was reduced. At one level, this may be accounted for by an increase in string familiarity, but a more general explanation is provided by a change in criteria setting. While there is

some question of how valid applying random walk models to IR can be, criteria setting does seem to be a promising, if largely descriptive, account of how task properties may moderate IR. Further there is wealth of literature indicating that criteria setting explain performance changes.

Criteria setting can of course be under conscious control. In agreement with this, participants reported firstly, they were aware when instances did repeat and secondly, realised that training targets were appearing in an invariant position. Together this supports the claim that IR is a conscious strategic choice (Haider & Frensch, 1999a), at least at the point of employment. Interestingly, there is some indication that response criteria are positioned by stimulus probability (Treisman & Williams, 1984), However, Ratcliff (1981) found that response boundaries were not changed by the probability of stimulus type. The concept of stimulus probability is therefore an interesting question when applied to IR and so will form the focus of the final experiment of the thesis.

## CONCLUSION

The study set out to test the general hypothesis that intrinsic task attributes may moderate IR, particularly;

- i. separability of relevant information from irrelevant,
- ii. overall task difficulty and
- iii. the number of unique instances presented.

The results support the hypothesis, in particular, reducing the number of unique string instances seems to increase IR. There is some indication that separability and task difficulty also contribute, to a lesser extent, to the effect.

The present study has provided evidence that intrinsic task attributes may moderate IR. Before this study there was only an indication that extrinsic task variables such as task-instructions moderated the effect. The very robust effect found with the AAT, also seems in part to result from one or more of the attributes tested here. This further indicates the need to investigate IR within a range of tasks. The following section of the study is concerned with an investigation into individual differences and how these impact on the IR strategy.

## 6.3 Individual Differences and Self-Report

### 6.3.1 RATIONALE

In the experiments reported here and in the earlier work of Haider and Frensch (1996, 1999a, 1999b, 2002), it becomes clear in each study that while generally participants did display Information Reduction (IR) over practice, this was not true of every individual. In each experiment there appeared to be a sub-set of participants who did not reduce. Whilst it is true that individual differences are expected in many types of psychology experiments, the repeated occurrence of a group of non-reducers is interesting. Haider and Frensch (1999b) noted this in a study investigating eye movement and IR. They differentiated between ‘Reducers’ and ‘Non-Reducers’ (by a median split of the regression score they used to measure IR), and found that non-reducers tended to scan more of the whole stimulus string, whilst reducers fixated only on the relevant segment.

A later investigation of the Power Law of Practice by Haider and Frensch (2002) also found evidence of reducing and non-reducing groups. The main focus of their study was to report that instead of the learning function following a continuous curve, as the power law predicts, the data from some participants showed discontinuities in learning. Haider and Frensch suggested that these discontinuities indicated the sudden adoption of a learned strategy. They used data from their IR experiments to support this and found *mostly* that participants displaying the discontinuity demonstrated IR with both a reduction in the string length effect and an increase in post-triplet errors during the test phase. However, participants who did not appear to reduce on the usual measures of IR also did not demonstrate the sudden drop in reaction times over blocks. This further supports the

possibility that there may be a sub-set of individuals who do not reduce the processing of irrelevant items during practice at a task.

The two experiments described above found, using a number of measures of IR, that some participants reduce while others do not. It would seem worthwhile therefore, to distinguish between these two groups. Such a distinction may be approached by just forming a taxonomy from the data, such as the median split suggested earlier. Perhaps though it would be more interesting to determine some other measure of individual difference that would predict whether a participant was likely to reduce or not. Most usefully this would take the form of some psychometric variable that could be measured pre or post the experiment.

### **6.3.2 CATEGORISING THE DATA**

Before any alternate measure of individual difference can be applied, some criterion for ascribing participants as reducers or non-reducers within the IR experiment has to be formulated. As already indicated, a median split of the data may prove sufficient. The split could involve the magnitude of the string length effect. The string length effect in this case was calculated as a regression slope using the length of the stimulus string as a predictor of Reaction Time (RT). As participants reduce, item length becomes a poor predictor and the slope value declines. Thus, someone who had reduced to processing the task-relevant items during a training block will produce a smaller regression slope value than someone who had not. This difference is most likely to be greatest at the last practice block. The coefficient data for this block therefore could be the focus of the median split

Alternate measures that could be taken from the data are the difference in coefficient value between the first and last practice block. Participants that show little reduction in the string-length effect will have coefficients that change little across blocks, whereas participants that reduce well, will show a decline in the coefficient across these blocks. If the value in the last practice block is subtracted from the first, the result will be a continuous variable; reducers having a higher value than non-reducers. This value can again be subjected to a median split, dividing into reducer and non-reducer categories.

The error rate in the final test block should also indicate the level of reduction that had occurred per participant. The test block differed in that for some of the strings the target appeared outside its usual position. The rationale was that if a reduction strategy was adopted during the practice trials, the test block strings with the target in the novel position would be incorrectly verified, leading to an increase in errors for this block compared to previous blocks. Overall, this provided evidence that IR had occurred in the experiment as a whole, but may also be used to determine if a single participant had reduced compared to another.

Perhaps a slightly less direct measure is reaction time (RT). As the irrelevant part of the stimulus string becomes ignored not only will there be less of a string length effect on RT, but overall mean RT for a block should reduce. The overall measure is less direct and confounded by any reduction in RT due to practice effects, but as this would be the case for each participant it may still be a useful indicator of IR. A way of isolating the difference in RT due to IR could be to look at the change in RT from the last practice block to the final test block. The rationale here is that non-reducers will have continued to search the whole

string in the last practice block and will continue to do this for the test block. Subsequently for each participant, their mean RT for these two blocks should differ very little. Reducers in contrast, will by the last practice block only process the relevant segment of the string, leading to an improvement in RT.

However, this strategy will not work for the test block and the feedback should prompt reducers to start searching the whole string again, increasing RT. Thus, non-reducers should show little change in RT between the two last blocks, while reducers should show more difference. A possibility to further enhance RT as a measure to help categories the level of reduction, is to look at the same difference suggested above, but restrict the analysis to only the longest seven item strings. It is these longest strings where the difference in RT for reducers and non-reducers will probably be most pronounced. This method is perhaps more likely to distinguish all the reducers than looking for RT discontinuities, which assume IR must occur as a sudden step-change in processing (Haider & Frensch, 2002)

The methods above suggest ways of determining the level of IR for a participant from the experimental data. Using the error rates and differences between blocks in coefficients and RT will result in a continuous variable suitable for analysis. This is valid, as IR may occur at variable degrees across individuals and practice. It is possible that the strategy may be used at different levels; sometimes checking the whole of the string or giving the irrelevant segment attenuated attention, with this attenuation increasing over blocks. On the other hand, a categorical variable may be useful. IR may occur acutely for some and they may exclusively use this strategy. Haider and Frensch (2002) provide evidence that a sudden



switch in strategy can occur, resulting in a discontinuity of RT over practice blocks. Because of this, it is worth looking at the data calculated both as a continuum and as categorical median split.

### **6.3.3 INDIVIDUAL DIFFERENCES MEASURES**

While discriminating the data in these ways is useful in itself for the overall analysis and investigation of IR, the second purpose here is produce some form which can be compared to other personality variables that may indicate the likelihood that an individual will utilise an IR strategy. To this end the data calculated from the above methods are used as a variable to investigate if IR can be predicted from individual psychometrics. Of further interest is the reliability of self-report data as an indicator of actual behaviour during the experiment.

There are of course a great many potential ability and personality variables that could be considered. Such a pursuit could constitute an entirely separate thesis. For this reason just two such measures were selected as an initial investigation, these were anxiety level and cognitive ability. The reasons why each of these was chosen and the selection of particular tests is explained within each following section.

#### **ANXIETY**

There is a great deal of literature to suggest that anxiety effects performance at a task. In a series of experiments Eysenck (1979,1985 & 1988) found participants high in Trait/State anxiety had lower attention or working memory capacity than those with lower measured anxiety. The reason suggested for this was that highly anxious individuals were subject to

task-irrelevant thoughts, such as worry or self-concern over performance. In a similar way those high in trait anxiety have also been found to be high in distractibility, see Eysenck (1982) for a review. Most of the studies on distractibility look at performance at a task in the presence of task-irrelevant stimuli. Just as highly anxious people allow task-irrelevant thoughts to intrude, it seems extraneous stimuli also intrude, using up resource for the main task with subsequent deterioration of performance.

Studies of attentional selectivity (Easterbrook, 1959) have shown a decrement in task performance following induced anxiety. The reduction in performance has been explained as a reduction in the breadth of attention as anxiety rises, with fewer irrelevant and finally relevant stimulus cues attended as arousal reduces ability to cope. Experimentally this view has been supported by studies using a dual task paradigm (Weltman, Smith & Egstrom, 1971 and Eysenck 1982). It was found that while a primary task was unaffected by increased arousal, a secondary task did show an inferior level of performance. Thus, this leads to two conflicting predictions regarding anxiety and IR, high anxiety could lead to either increased or decreased reduction.

Overall though, the suggestion is that resources are withdrawn from a secondary task as anxiety is increased. That is, anxiety causes a perceptual narrowing and increased allocation of resources to the main task. This is done in an attempt to compensate for the resource demands of anxiety itself. Murray and Janelle (2003) also found evidence of attentional narrowing for a High Anxious (HA) group during a dual task simulation. Firstly, they found reduced performance in a secondary response-time task for the HA group over blocks whereas conversely practice produced quicker responses for a low anxious (LA)

group. Secondly, HA individuals displayed inferior search patterns for the secondary task, with less efficient saccades to the peripheral stimuli.

In general this suggests that while highly anxious individuals may be very distractible, the reduction in resource 'space' resulting from anxiety causes attentional narrowing in an attempt to compensate for such deficits. It would appear that this narrowing can take the form both of increased allocation of processing towards what is regarded as a primary task and also a narrowing of the attentional spotlight (Eriksen, 1985, 1986; Murray & Janelle, 2003). Regarding information reduction in this light, it would seem likely that the anxious individual with reduced resources may well show a predisposition to have a narrowed field of attention and also direct this attention to what is found to be the relevant part of the task.

So it could be hypothesised that such individuals will display IR to a greater extent than low anxiety individuals. Of course there is also some evidence above that HA individuals are very distractible and show inefficient search strategies. Also LA individuals may have greater spare resource to form alternate strategies such as IR. It does seem though, that HA groups allocate their attention to a specific area. This view is partly supported by increased IR when participants are put under speed stress (Haider & Frensch, 1999a). Consequently a measure of anxiety would appear to be a good candidate to further the understanding of individual differences in IR. Whilst it is an empirical question as to the exact effect of anxiety upon IR, the prediction is that HA individuals will reduce more as they try to focus resources and attenuate secondary or irrelevant or information.

In the current investigation anxiety was not manipulated directly, previously ‘Reducer’ and ‘Non-Reducer’ groups appeared to form in each experiment without this manipulation. Rather ‘Trait’ anxiety was measured by a questionnaire sent to participants to fill in before the day of the study and ‘State’ anxiety was measured just before the commencement of the study session. These two measures were taken as there is evidence that State and Trait anxiety are separable dimension of emotion (Spielberger, 1969).

State anxiety is transitory and depends on the demands of the moment and predisposition to anxiety. This predisposition can be regarded as trait anxiety, simply how anxious the individual is, as a long-standing general condition. This distinction is interesting for the purpose of segregating groups into reducers and non-reducers as it allows measurement of both the individuals’ normal affect and also how anxious they become before the test. Because of this, Spielberger’s (1969) State-Trait Anxiety Inventory was selected. This test is both highly correlated with other anxiety tests (such as the Multiple Affect Adjective Check List, MAACL) and was easy to administer as a postal questionnaire, so overall was ideal as a first investigation into IR and anxiety.

### *Method*

For all 48 participants of IR Experiment 4, a copy of the STAI trait questionnaire was posted out, with the instructions to fill it in to reflect how they generally felt. The instructions also asked that the form be completed at least one day before the experiment. The questionnaires were returned when participants came to the study session. At this time, following reading and agreement to an informed consent form, the STAI state questionnaire was filled out, on this occasion participants were instructed to fill the form out to reflect

how they felt at that moment. The form was then removed and the study commenced. All the data was collated and the questionnaires scored following Spielberger's (1969) instructions.

### *Cognitive Ability (IQ)*

As a second line of inquiry, there is evidence that cognitive ability or cognitive style may moderate the development and selection of strategies. On initial reading of the literature the concept of 'cognitive style' appears promising as a way to differentiate those likely to reduce from those that do not. Overall, the research follows the line of testing dichotomies in the way people 'choose' to process information. Perhaps the classic example of this is the *Holist/Serialist* distinction (Pask & Scott, 1972). Their experiments indicated that there were two identifiable individual differences in strategy usage. The holist seeks information to obtain a global picture of the domain, whilst the serialist uses a step-by-step approach seeking information about each domain element before moving on. The concept of different cognitive styles may differentiate between those that select an IR strategy and those that do not. For instance, holists may view the string as a whole while serialists are more likely to distinguish the separate relevant and irrelevant segments.

However, as suggested by Roberts and Newton (2001) and Tiedemann, (1989), there are many such dichotomies in the literature. Such as,

- i. *Visualisers/Verbalisers,*
- ii. *Field Dependant/Field independent,*
- iii. *Assimilators/Explorers,*

- iv. *Adaptors/Innovators and*
- v. *Convergent/Divergent Thinkers*

The above list whilst categorising a number of individual differences in strategy use, also makes it problematic to decide which if any of the many 'styles' may predict IR. Indeed the Reducer/Non-Reducer distinction could be offered as an additional dichotomy. It may be more fruitful to look not at different styles, but rather the abilities that cause people to select a particular strategy.

Tiedemann (1989), further remarks that the dichotomies present in the Cognitive Styles literature are in fact just describing behaviour, rather than explaining it. Additionally he suggests that most 'styles' are validated against competence tests and so most cognitive styles can be integrated into Guilford's (1980) Structure of intellect Model. Thus, a cognitive style may just be a re-expression, or result of some underlying individual difference in either central or specific intelligence.

This view is supported by Roberts (2001), who suggests that it is necessary to understand how the underlying abilities of an individual impact on strategy selection. In a study using a compass navigation task, Roberts *et al* (1997) found that participants high in spatial ability tended to use a non-spatial canceling strategy, while those with low spatial ability paradoxically continued with a spatial strategy to follow the compass point directions. Such a result was very difficult to explain as the verbaliser/visualiser cognitive style concept.

Roberts wished to test this concept thoroughly, and the results found it lacking. While 'cognitive style' cannot explain these findings, an ability account fairs somewhat better. The crux of the explanation was that participants high in spatial ability are better able to develop the more efficient cancellation strategy. Whereas, those low in this ability struggle with the task and so do not have the resources to 'discover' the better method. So, strategy selection in this task seems mediated by individual differences in spatial ability, though not as a cognitive style. This is reasonable evidence that a specific intellectual ability may underlie strategy selection in this case.

The role of more central intelligence is shown in a study by Schunn (2001). The study demonstrated that the facility to adaptively apply strategies to meet changing task demands is primarily associated with reasoning ability and working memory capacity. During tests with the Kanfer-Ackerman Air Traffic Controller task (KA-ATC; Kanfer & Ackerman, 1989), Schunn (2001) found ability measures of spatial procedural learning and spatial inductive reasoning were predictors of strategy adaptivity under changing base rates of success. Both these measures were part of the Cognitive Abilities Measurement (CAM) tool and overall suggest adaptivity was related to reasoning ability.

It would appear, that some measure of central intelligence might therefore be a viable predictor of strategy selection in general. More specifically, it may provide some differentiation between those that adopt IR and those that do not discover this strategy. In a

personal communication with Christian Schunn (4<sup>th</sup> Nov. 2003)<sup>1</sup>, he felt that a Raven's type Matrix reasoning task would be a good central measure of intelligence for this purpose.

So, for the initial investigation into individual differences and IR, the matrix-reasoning sub-test of the Wechsler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999) was chosen. This test is suitable for all age groups and takes approximately five to fifteen minutes to administer. This is a particular advantage, as participants will have already completed quite a long and repetitive study session and a long intelligence test would be counterproductive.

### *Method*

Twenty four participants (six from each group) from Experiment 4 were given the test. It immediately followed the target search task and questionnaire. Although timed, the test was self-paced.

Participants were not told directly it was an intelligence test (as this sub-test on its own does not warrant such a description), instead they were told a comparison between their performance on the target search task and ability at puzzles was required. Participants were told they could stop at any time and could skip puzzles they could not complete. (some examples of the matrix task are in appendix 3). Participants were not informed of their score. All the test sheets were scored later using a small program to compute the results.

---

<sup>1</sup> Schunn suggested we may be finding strategic accuracy bias, though, rather than adaptivity differences. Bob Siegler found that some children generally prefer to use backup strategies (like counting on fingers) every time rather than retrieve because they want to be sure they are correct. He called them perfectionists. He found this in reading and in math, although it tended not to correlate across domains. Overall, they were quite accurate. He suggested we might ask people about how much they value speed and accuracy to see how this relates to their groupings.



Apart from the overall score, results for items requiring spatial rotation and items involving pattern matching were computed separately.

### *Self-Report*

Haider and Frensch (1999a) found that IR increased under speeded compared to accuracy instructions. This they took as an indication that IR was under conscious control. However, whether Participants are generally aware of the underlying rule that allows IR in the tasks reported here has not been fully reported, especially for the target search task we have used. To address this, at the end of the target search trials in Experiment 4, a questionnaire was given with items intended to 'draw out' knowledge of the rule that relegates some of the stimulus information redundant. That is, if a target was present it was always in the second position from left.

Not only is it interesting to determine if participants explicitly learned the rule and reported it, but also it is interesting to discover if self-report does predict behaviour at the task. It may be that people do know the rule explicitly but do not always choose to employ it, or conversely people may sometimes know the rule implicitly and this affects behaviour, but it is not reported. The questionnaire was given to all the 48 participants in the study and an example of the questions can be found in appendix 2.

## **6.3.4 RESULTS**

### **ANXIETY**

The STAI scores from the Y1 State anxiety questionnaire (how anxious you feel at this moment) and the Y2 Trait anxiety questionnaire (how much anxiety you generally

experience) were subjected to a 2-WAY ANOVA, Condition (1-4, between subjects) \* Anxiety Test(State-Trait, within subjects). There was no main effect of Condition, but there was a main effect of Anxiety Test  $F(1,44)=30.854$ ,  $p < 0.0001$ , score means show State anxiety to be reliably lower than Trait anxiety. Both scales measure on a four point scale and have been shown to correlate with each other. However, while the construct of anxiety formulated by Speilberger (1969) underlies both scales, the questions are phrased differently in Y1 and Y2. So this does not suggests participants were less anxious than normal during the test. There was no Condition \* Anxiety Test interaction.

using Y1 & Y2 together as part of a randomised controlled trial

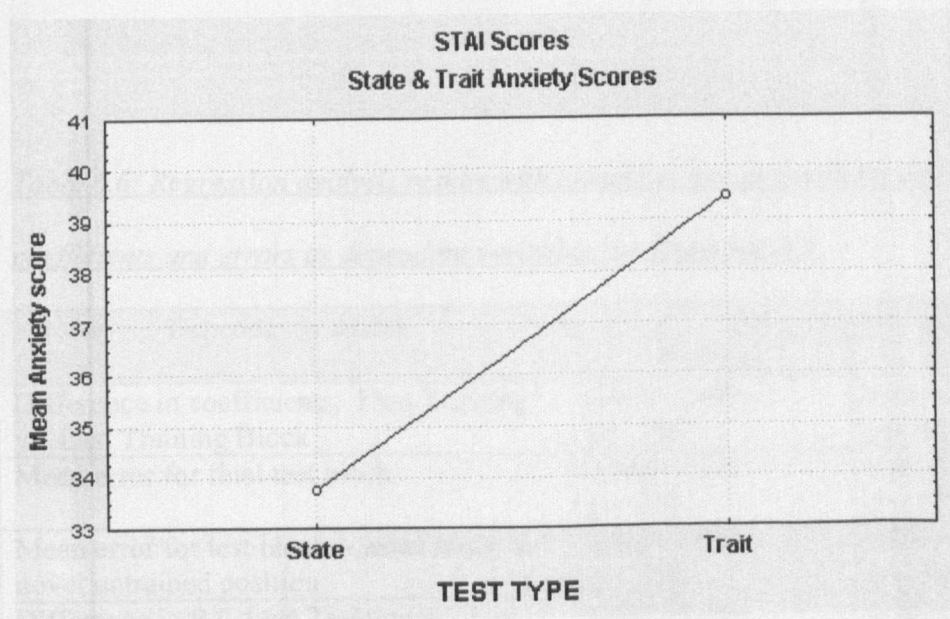


Figure 6.8: Mean State and Trait anxiety scores for the STAI questionnaires. Scores can potentially range from 20 to 80 points in the inventory; for Experiment 4.

Overall the results suggest that participants were not particularly anxious about the study scoring less than 34 out of a potential 80 points. Also, there were no baseline difference in anxiety between groups. Multiple regression show STAI Trait anxiety to be a significant

predictor of State anxiety  $t(46)=5.851$ ,  $p< 0.0001$ , suggesting that while state anxiety may differ in level from trait anxiety, there is the expected interdependence between the two.

### Discrete variables

To determine if anxiety level as measured by the STAI predicted IR, the State and Trait scores were used as predictors in a linear regression analysis. The dependant measures were the continuous variables described in the rationale as possible indicators of IR. The results in Table 6.6 below, indicate that Anxiety as measured by the STAI did not predict scores on these variables. Neither Y1(State) or Y2 (Trait) are significant, neither were overall models using Y1 &Y2 together as part of a multiple regression analysis.

### Logistic regression analysis

Median splits for error scores and RT were used

### In novel position

*Table 6.6: Regression analysis results with anxiety scores as predictor and mean coefficients and errors as dependant variables, for Experiment 4.*

Dependant variable	Independent Variable	
	Y1 State	Y2 Trait
Difference in coefficients, First Training vs. Last Training Block	$t(46)= -.869$ $p= .389$	$t(46)= -.087$ $p= .931$
Mean error for final test block	$t(46)= -.478$ $p= .635$	$t(46)= -.422$ $p= .675$
Mean error for test block – when target in novel untrained position	$t(46)= -.906$ $p= .370$	$t(46)= -.972$ $p= .336$
Difference in RT, Last Training vs. Test Block	$t(46)= 1.254$ $p= .216$	$t(46)= .252$ $p= .802$
Difference in RT, Last Training vs. Test Block, For 7 item strings only	$t(46)= 1.316$ $p= .195$	$t(46)= .647$ $p= .521$

### Logistic regression analysis

Logistic regression was used to determine if the categorical variables (median splits of coefficients and error scores into ‘reducer’ and ‘non-reducer categories’) were predicted by anxiety scales Y1 & Y2. The results shown in Table 6.7 find that neither Y1 or Y2 provide a significant prediction of these variables, a model combining Y1 & Y2 again found no



significant prediction. This further supports the conclusion that the level anxiety as measured by the STAI does not predict IR values calculated in either the continuous or discrete variables.

*Table 6.7: Logistic regression results, anxiety scores as predictors of reducer and non-reducer categories, for Experiment 4.*

Dependant variable	Independent Variable	
	Y1 State	Y2 Trait
Median split of last practice block coefficients	t(46)= .078 p= .938	t(46)= .373 p= .710
Median split on difference between coefficients, first vs. last training blocks	t(46)= -.469 p= .641	t(46)= -.276 p= .783
Median split for errors in test block, target in novel position	t(46)= .352 p= .727	t(46)= .470 p= .640

#### COGNITIVE ABILITY (IQ)

As for the anxiety scores, the WASI scores were used as a predictor for the continuous IR dependant variables. Apart from the total WASI score, results for spatial and pattern-matching elements of the test were also used as part of the regression, as was the total time taken to complete the test. The results in Table 6.8, suggest that only scores on Pattern Matching items of the WASI predicted any of the dependant variables. Both overall errors in the test block and specifically errors during test where the target was outside its normal trained position were predicted by pattern matching scores. The correlation was negative, suggesting that those better at pattern matching missed fewer targets when it was in the formally redundant segment of the string. No other models were significant.

*Table 6.8: Regression analysis results with WASI scores as predictor and mean coefficients and errors as dependant variables, for Experiment 4.*

Dependant variable	Independent Variable			
	Total WASI	Spatial	Pattern Matching	Time
Difference in coefficients, First Training vs. Last Training Block	t(22)= -.783 p=.088	t(22)= .099 p=.922	t(22)= -.867 p=.395	t(22)= .46 p=.65
Mean error for test block	t(22)= -1.309 p=.204	t(22)= .236 p=.816	t(22)= -2.298 p=.0315*	t(22)= .621 p=.541
Mean error for test block – when target was in novel untrained position	t(22)= -.078 p=.293	t(22)= .371 p=.714	t(22)= -2.41 p=.0248*	t(22)= 1.156 p=.26
Difference in RT, Last Training vs. Test Block	t(22)= .732 p=.472	t(22)= .887 p=.384	t(22)= -.797 p=.434	t(22)= .949 p=.353
Difference in RT, Last Training vs. Test Block, For 7 item strings only	t(22)= -.212 p=.834	t(22)= .257 p=.799	t(22)= -1.34 p=.194	t(22)= .328 p=.746

Logistic regression was again used to investigate if the median split variables were predicted by any of the WASI scores. As indicated by Table 6.9 None of the intelligence scores provided significant prediction of these discrete variables, a model combining the three measures again found a model with no significant prediction.

*Table 6.9: Logistic regression results, WASI scores as predictors of reducer and non-reducer categories, for Experiment 4.*

Dependant variable	Independent Variable		
	Total WASI	Spatial	Pattern Matching
Median split of last practice block coefficients	t(22)= 1.018 p=.32	t(22)= .00 p=1.0	t(22)= 1.156 p=.26
Median split on difference between coefficients, first Vs. last practice blocks	t(22)= -.579 p=.568	t(22)= .291 p=.774	t(22)= -.427 p=.674
Median split for errors in test block, target in novel position	t(22)= .343 p=.734	t(22)= .936 p=.36	t(22)= -1.515 p=.144

## DISCRIMINANT FUNCTION ANALYSIS

For thoroughness, as a final test all the Anxiety STAI and intelligence WASI variables were entered into a Discriminant function analysis to determine if the groups generated by the median splits were discriminated by these scores. No significant model was found. This further supports the conclusion that overall neither the STAI or WASI scores predict the level of Information Reduction calculated in the various ways.

## SELF-REPORT

The data from the self-report questionnaires were classified into those that reported realising the rule (Target if present, always second item from left), and those that did not state any knowledge of the rule. To determine if those that report knowing the rule were significantly different in IR behaviour to those that do not express the rule, Chi-square was calculated for frequency levels both of Know/not know the rule \* Median split of final practice block coefficients, and also Know/not know the rule \* Median split of error rate for test block target in novel position.

The distribution of observed frequencies (see table 6.10a) for those did/not know the rule (Yes/No) and the reducers non-reducers as categorised by a median split of final practice block coefficients, differed significantly from an expected random distribution. Chi-square  $\chi^2(1) = 9.55, p < 0.01$ ,



Table 6.10a: Chi Square results for those that reported knowing the rule or not \* ‘reducer’, ‘non-reducer’ categories

MedianSplit * SelfReport			
	yes	no	Total
reducers	23	1	24
non-reducers	14	10	24
Total	37	11	48

For those participants who knew the rule the frequency for the *reducer* category (those grouped into a low final training block regression coefficient) was higher than that for the *non-reducer* category (those in the high-coefficient group), 23 vs. 14. Whereas, for those that did not report knowing the rule this, was reversed with 10 *non-reducers* not knowing the rule and only 1 *reducer* not reporting the rule.

Table 6.10b: Chi Square results for those that reported knowing the rule, thought they knew the rule or did not know it \* ‘reducer’, ‘non-reducer’ categories

MedianSplit * SelfReport2				
	yes	maybe	no	Total
reducers	18	5	1	24
non-reducers	7	7	10	24
Total	25	12	11	48

Further investigation by splitting the self-report data into those that knew the rule(Yes), those that were less sure(Maybe) and those that had no idea of the underlying rule(No), found a similar result (see table 6.10b). Chi-square  $df(2) = 12.537$ ,  $P < 0.01$  suggests more clearly that more participants that reported knowing the rule reduced during the study than those that did not report such knowledge.

*Table 6.10b: Chi Square results for those that reported knowing the rule or not \* 'reducer', 'non-reducer' error categories*

*Table 6.10c: Chi Square results for those that reported knowing the rule or not \* 'reducer', 'non-reducer' error categories*

The self-report data was also compared to a grouping of the error data from the final test block (Target in novel untrained position), with the data categorised as high or low errors, reflecting reducers or non-reducers respectively (see table 6.11a). Self-report split into Yes-knew rule or No-did-not know rule found a significant distribution difference from expected, Chi-square  $df(1) = 9.553$ ,  $P < 0.01$ , once again more of those that knew the rule fell into the reducer category than non-reducer. While this is reversed for those that did not know the rule.

*Table 6.11a: Chi Square results for those that reported knowing the rule or not \* 'reducer', 'non-reducer' error categories*

*Table 6.11a: Chi Square results for those that reported knowing the rule or not \* 'reducer', 'non-reducer' error categories*

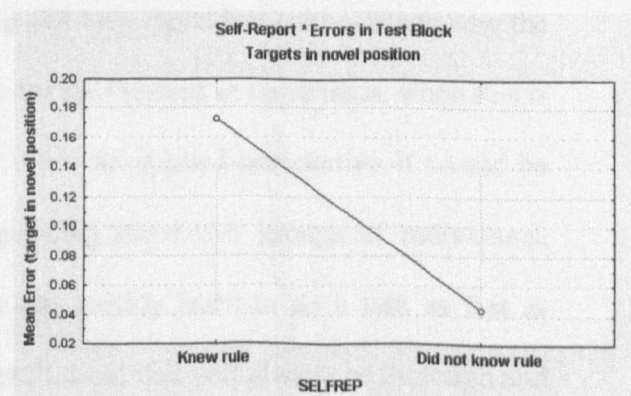
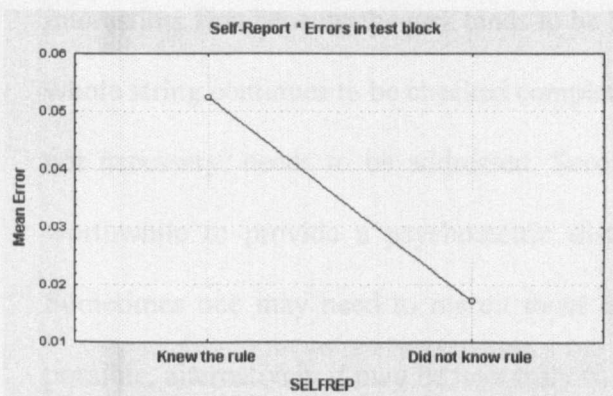
ErrorSplit * SelfReport			
	yes	no	Total
reducers	23	1	24
non-reducers	14	10	24
Total	37	11	48



Splitting self-report into yes-maybe-no once again repeats this pattern. Chi-Square  $df(2)=9.697$ ,  $p<0.01$ , overall suggesting that those reporting the rule were more likely to reduce during the experiment than those that did not report the rule.

*Table 6.11b: Chi Square results for those that reported knowing the rule, thought they knew the rule or did not know it \* 'reducer', 'non-reducer' error categories*

ErrorSplit * SelfReport2				
	yes	maybe	no	Total
reducers	15	8	1	24
non-reducers	10	4	10	24
Total	25	12	11	48



*Figure 6.9: Mean overall Error rate and Errors for novel target positions in the final test block, participants that reported the rule or not, from Experiment 4.*

As a final test the continuous mean error rates were subjected to ANOVA with Self-Report as the grouping variable. As indicated in figure 6.9, there was a significant main effect of

Self-Report for the level of Error,  $F(1,46)= 4.71$  ,  $p=.035$  for the mean overall error rate in the test block.

Also there was a significant main effect of Self-Report for Error,  $F(1,46)= 4.566$ ,  $p= .038$ , for the mean error rate in the test block when targets were in a novel untrained position. There were significantly more errors when the rule was reported than when it was not reported. This suggests that people reduced more when they were consciously aware of the rule.

### **6.3.5 DISCUSSION**

Previous experiments have found that consistently there was a sub-set of individuals that did not use an IR strategy (Haider & Frensch, 2002; Thesis studies 1-3). This was interesting, first because the task tends to be long and very repetitive, so the reason why the whole string continues to be checked completely for six hundred or more trials, when this is not necessary, needs to be addressed. Second, from an applied perspective it would be worthwhile to provide a psychometric distinguishing these two groups of individuals. Sometimes one may need to recruit those that can quickly learn to do a task as fast as possible, alternatively it may be necessary to recruit those that will always be thorough and check every component no matter how routine. Subsequently, the aim here was to provide a valid method of dividing participant data into reducers and non-reducers and then provide a metric to distinguish between these two groups.

## THE EFFECT OF ANXIETY UPON IR

Neither measured State or Trait anxiety was found to predict whether participants reduced or not in Experiment 4 of this chapter. However, there are a number of reasons why this may have been the case. The task and experimental situation appeared not to be very stressful; state anxiety measured just at the start of the experiment averaged under 34 out of a possible 80. State anxiety was also significantly lower than trait anxiety. So high levels of anxiety were unlikely to be a factor within the experiment. It is possible that more extreme groups would be needed to find an effect. It was decided not to manipulate anxiety since non-reducers were found in previous experiments where this manipulation had not taken place.

Another problem was the level of anxiety was not measured at the end of the task session. Again this could have reduced the chance of finding an effect of anxiety. The reason for not taking this second measure was that the questionnaire was not really suitable for repeat administration over so short a time scale. There is also some question over how well a median split identifies reducers. Admittedly, this is a crude measure, where individuals close to the division may not easily fit into either category. Thus, there were a number of problems relating anxiety to reduction in this study.

However, even with these shortcomings in mind, the aim was to have a psychometric that would predict IR, and distinguish between potential reducers and non-reducers. So, it seems safe to conclude that trait anxiety as measured by the STAI at least, was not a useful measure for this purpose. That is, overall anxiety at the start of the task did not predict whether a participant would reduce or not. As anxiety was not an experimental

manipulation, it is too early to rule out the possibility that higher states of anxiety may moderate IR. There is significant evidence in the literature that attention is moderated by anxiety and IR is an attentional strategy. Therefore, a possible area for future research could be an investigation into IR under anxiety inducing situations or tasks.

## THE EFFECT OF COGNITIVE ABILITY UPON IR

The cognitive ability measure employed was also largely unsuccessful in distinguishing between the reducers/non-reducers. The matrices correlate with verbal ability and so are a reasonable measure of central intelligence. However, it seems that general intellectual ability, as measured by the scales, did not predict strategy use in this case for this particular task. There was some indication that the pattern-matching scores did predict IR, but considering the number of comparisons that found no result, this finding is tenuous to say the least. If valid though, it does suggest that those that are efficient in focusing on detail are less likely to reduce. This could suggest that those participants that do not reduce, behave in this way because they are very thorough in everything they do. For these individuals, IR may only occur if the cost of not reducing is great, such as when the complexity of the task is high or there is some speed criteria. This does tie in with Schunn's suggestion that IR may be some sort of accuracy bias, what Seigler (1989) calls perfectionists. Some test of this thoroughness might prove a better dimension to distinguish reducers from non-reducers. Maybe asking individuals how much they value speed at a task versus accuracy.

## SELF-REPORT DATA AND IR

The Self-report data was interesting as there was a question over explicit knowledge of the rule. That is, do people have to know the rule explicitly to apply it? Alternately, do people know the rule and choose not to use it.? It would seem from the self-report analysis that generally, when the rule is realised it is used. Those expressing explicit knowledge of the rule do reduce more than those that do not. Reduction in this case, measured by variables such as reduced string length effect and increased errors when the target appears in the formally redundant area of the stimulus.

There is still the possibility that knowledge of the rule passes through an implicit stage, before it becomes explicit and indeed the questionnaires did suggest that some participants only expressed the rule following some prompt. However, the results suggested overall that IR was a conscious strategy and the rule once known was applied in the reduction experiment of the present study. Of course, there is research to suggest the rule may not always be applied even if known. Haider and Frensch (1999a) found the strategy was only applied when there were speed instructions and we also found attenuated use of the rule when the advantage over not reducing was not that great. So employment of the 'rule' may still be dependant on task demands

The correspondence between the self-report data and the median-split of regression coefficients for the last training block, suggests the latter to be a valid variable on which to isolate reducers/non-reducers in the data. The same is true for error rates in the test block, again a median split of the data seems to distinguish between the two groups. Such a

distinction will be worthwhile for further research if the effect of task dimensions on reducers is to be investigated separately from the effect on the overall population to reduce.

## SUMMARY

Further investigation is needed to determine the role of anxiety in IR strategy selection. Further work is also needed to find a metric that will predict the probability of such a strategy choice. There is some suggestion that a measure of thoroughness or perfectionism may prove a possibility. Additionally, a more in-depth form of self-report such as 'think aloud' protocol analysis to investigate individual differences in learning may prove useful.

### *Summary of Study 3(Experiment 4 & Individual Differences)*

This study found that anxiety level at the start of the task did not influence reduction. Unfortunately, neither did the cognitive ability measures really distinguish between reducers and non-reducers. This was not surprising considering the number of individual and personality differences that potentially may underlie performance. Indeed, such an investigation is a separate line of research of its own. For this reason the investigation of individual differences and IR will be left at this point for future research that is beyond the scope of this thesis. To end this on a positive note, the self-report data did seem to indicate that those that reduced were conscious of the rule and this did effect the level of reduction.

The experimental data also support the hypothesis that certain intrinsic task attributes moderate IR. Particularly, instance repetition apparently increased reduction, though task difficulty and string segmentation also seemed to contribute to the effect. An important finding since most of what is known about IR comes from the AAT, which had a specific

set of attributes. The final study of the thesis investigates yet another attribute, the regularity of relevant and redundant information.

## **CHAPTER 7: STUDY 4**

### **The Regularity of Source Redundancy**

*The aim in running this study was to determine if it was necessary for the irrelevant segment of the stimulus to be consistently redundant for information reduction (IR) to develop.*



## 7.1 INTRODUCTION

The previous chapter concerned task attributes that might facilitate IR. It would seem that reduction increased following a number of manipulations, particularly when the number of unique task instances was reduced. This would suggest that ‘real world’ tasks, with this element of repetition involved, might be more prone to the strategic use of IR. The studies previously run suggest IR is open to a number of moderating task attributes; however, these moderators may not just facilitate, but also inhibit reduction. For example, instructions to be as accurate as possible inhibit reduction when compared to speeded conditions (Haider & Frensch, 1999a). In a later study, again using the AAT, triplet placement also changed the level of reduction. While placing the triplet at the end of a string did not moderate reduction, partly ruling out a fatigue explanation of IR, inconsistent triplet placement did (Haider & Frensch, 1999b). Specifically, when triplet position was randomly moved from first to last position in a string across trials, while IR occurred it was reduced. This suggests there was less reduction when the probability that relevant information would be at one location was lowered.

Apart from this exception, the pure experimental tasks described so far have held the redundancy constant. For example, the triplet was always relevant in the AAT, while the post-triplet letters were irrelevant. Even when triplet position was randomised, it always remained the relevant information source. Of course, during the test phases it was common for this regularity to change. In this latter phase, the formerly redundant segments become relevant, with targets occurring in untrained novel string positions. In all cases this caused the IR effect to begin to disappear; participants began once again to scan the whole string. This change in strategy was prompted, it seems, by the error feedback after each trial. Participants begin to switch back to processing the entire string

when they realised the IR strategy was no longer entirely successful. This demonstrates that IR occurred when the redundancy was always constant, but also that participants may have chosen to drop this strategy when it was apparent this consistency was no longer true. These points raise interesting questions over how consistent the redundancy has to be over task experience before an IR strategy forms and is used.

If IR only occurs if redundant items consistently remain entirely task-irrelevant, this limits the scope of tasks where it may develop. One can think of a number of tasks where some information is almost always redundant – but sometimes not. Indeed, this is why checklists are used within complex tasks. Often items or gauges that rarely realise a problem are ignored without such checking procedures. Thus, real world experience would suggest IR does occur, even when information redundancy does not always remain regular.

For instance, the Three Mile Island incident described in the literature review. On that occasion, gauges were ignored because they were not always necessary for the running of the plant. Though, they were sometimes relevant and became particularly so during the plant failure. The number of real skills that hold redundancy at 100% regularity are therefore likely to be constrained, so the ‘tolerance’ of IR to variations in the underlying redundancy rule is interesting. Some robustness is required, especially to accept IR as a significant part of everyday skilled performance.

### *Criteria setting and probability manipulations*

Random walk modelling (RWM) has been given some interest in this thesis as there is some evidence that the criteria setting that effects random walk processes may be associated with a similar effect in IR (Haider & Frensch, 1999a). That, is under more

liberal response criteria less information may be accumulated during a random walk (RW) decision process and this would be associated with reduced processing of task-irrelevant information. As expounded earlier, this is not suggesting that RW processes explain IR, they are two separable processes, rather the idea of criteria setting has some power in explaining both changes. So, the concept of random walk and boundary criteria seem a useful way to isolate manipulations that may direct IR. Of interest here is the change brought about by stimulus probability.

Ratcliff (1981) attempted model fits to data where probability was varied. The data came from a study using two alternate forced choice RT tasks. The experimental task was to judge if two letter strings were the same or different. The probability manipulation varied the number of same and different sets that occurred in a session. There were four probability conditions, 20-80%, 40-60%, 60-40% and 80-20% same-different trials respectively. The idea was that decisions would be easier at each extreme of the same-different probabilities, leading to faster response times. The main thing noted about the fit of the model to the data was that the bias manipulation was not strong enough to shift the boundary conditions. However, what did change, for the 80S-20D (S = Same, D = different) condition only, was the starting point for the information search, which in the model was 10% nearer the *same* boundary. This corresponded to a decreased reaction time for that condition.

For the 80S-20D condition, one could propose that less information was still sought before a response, as there was now a bias in the starting point of information search. A bias that changed the amount of information accumulated before the 'same' judgement was made. Treisman and Williams (1984) have also proposed that signal probability and

past experience will effect criteria setting. The effect of stimulus reliability within IR is an interesting question then.

At this point there is little evidence that strongly indicates probability manipulations will affect criteria setting and subsequently IR. There was some indication though, that for any change in information processing under the probability conditions mentioned, there needed to be at least an 80% bias. This may be an important threshold to bear in mind.

#### *Automated systems and information reliability*

The probability bias Ratcliff (1981) describes changes the likelihood that two sets of letter strings matched or not; within the IR paradigm this is of less interest. The concern is not how many correct vs. incorrect strings are needed to promote IR, but rather how reliable an information source needs to be before other sources are ignored. One area of investigation directly concerned with this is research focused upon automated warning systems. In this highly applied field, the reliability of warning systems is often a task variable used to determine when the effectiveness of a warning system fails. The reason for this interest is that no warning system can be 100% reliable under all conditions. So it is necessary to find out if a less than optimal system is better than none at all and under which conditions this applies. On pure performance measures it has been found that automated alerting systems that fall below 70-80% reliability tend to cause a decrement in performance. This was mainly because practitioners are aware of the sub-optimal information, but choose to depend on the automation to lower work load (Wickens & Dixon, 2005; Xu, Wickens & Rantanen, 2004).

Contrary to this, Bliss and Acton (2003) found using driving simulations, that reactions to a dashboard alarm declined significantly when reliability fell below 75%. They suggested the cry-wolf phenomenon as a reason for ignoring the alarm when it was often unreliable. Similarly, Lee and See (2004) found that mistrust in automated alerting systems increased rapidly as reliability dropped. The absolute level this occurred was found to be highly system and context dependant, ranging from 90% to 60%. So under some conditions low automation reliability can cause mistrust to form, and subsequently less reliance on this source. However, there is some evidence that under high work loads or high initial reliability in the system, sub-optimal automation is still relied upon (Lee and See, 2004; Wickens & Dixon, 2005; Wickens, Gempler & Morpew, 2000)

As far as IR applies to this, it would seem likely that people are sensitive to the reliability of an information source. However it also seems that this level of reliability does not need to be 100% for reliance on a particular information source to develop. It would appear people still depend or trust a source of information when its reliability as a predictor is as low as 60%, and trust certainly persists at 90% levels. Whether reduction strategies behave in this way is an empirical question. The research above seems to suggest IR might still occur as long as information remains redundant fairly consistently. That is, the regularity of the redundancy does not fall too much below 90%

### *IR and heuristics*

There is a great deal of literature that indicates people do use sub-optimal strategies, utilising reduced levels of information to make decisions (Kahneman, Slovic & Tversky, 1982). Rules of thumb or heuristics to name these strategy types are often used to perform tasks at an *acceptable* level. The key element that ties heuristics to IR strategies is that both proceed without using all the available information. In the case of

heuristics, the development of such methods is thought to follow either environmental or processing constraints (Gigerenzer & Selten, 2001). These include the need to perform the task quicker or reduce a complex task to a manageable level.

Heuristics are often regarded as strategies that lead to decision solutions. In essence though, they are simplified methods that cannot always be successful in all contexts, i.e. they do not guarantee success. Success rate alone though, does not define a heuristic method, since in real environments there are few strategies that will result in 100% correct performance. Rather, it is the level of simplification that defines the heuristic strategy. Generally, heuristics simplify the approach to a problem and make complex information easier to deal with. For example, car parking spaces are usually easier to find further away from the entrance to a supermarket. This is not always so, but generally makes finding a parking space easier than searching the rows systematically (which would be closer to an algorithmic solution). Hutchinson & Gigerenzer (2005) hypothesise that much of human decision making can be described by these simple 'rule of thumb' processes (heuristics).

Others have drawn links between human and animal research on this matter, suggesting that blocking can mean that multiple redundant cues may not all be used (Shettleworth, 2005). Also animal behaviour may depend on rules that while usually effective are not always optimal (Stolarz-Fantino & Fantino, 2005). The proposal is that for both humans and animals these heuristics have developed to achieve speed by seeking or using as few environmental cues as prudent (Gigerenzer & Selten, 2001). Generalised to IR, there would seem to be a pre-disposition to develop strategies that use as few relevant information sources (or those that are most efficient) as possible and ignore any redundancies.

Exploring this theme a little further, Gigerenzer and Selten (2001) have proposed four classes of heuristics,

- (a) Those that rely on recognition,
- (b) Those that use the first encountered cue that indicates an unambiguous choice,
- (c) Heuristics that combine a small number of cues to make a categorical decision and
- (d) Heuristics that stop a sequential search after encountering only a small number of alternatives.

Class (a), recognition heuristics, are a sort of ignorance-based decision, in that no information is sought. Rather only recognition processes are used to make a choice. Within the IR experiments described so far, it did seem that recognising the strings served to moderate reduction. That is, people seemed to ignore the redundancy more if they could remember the strings. Class (b), the first encountered cue that indicates a choice, again parallels the IR research, Green and Wright (2003) found that people reduced to the first relevant information encountered. Also, in the other tasks here and in the literature the relevant segment is encountered first and provides an unambiguous choice during training. The combining of a sub-set of cues to form a decision (c) has not been approached within IR yet and (d) does not really apply as the string items are not really alternatives in the sense described by Gigerenzer and Selten (2001).

IR and heuristics do seem to share some common principles. If this is the case, IR may also occur when it provides sub-optimal performance. In other words, people may still use a reduction strategy even if it does not always provide the correct answer. If people do apply IR in a heuristic fashion though, the next question is at what point a rule such as 'ignore the task-irrelevant post-triplet items' would be abandoned. That is, would

there be sensitivity to rule reliability that determines use. The 'automation' research already discussed would suggest this is possible and there is additional evidence that people are adaptive to success rates.

Schunn and Reder (2001) have investigated strategy use in the Kanfer-Ackerman Air Traffic Controller Task (KA-ATC or ATC). Within the task they manipulated the proportion of 'heavy' aircraft there were to land. Importantly, there were two runways, a short and a long. One of the problems faced in this speeded task was that heavy planes could not land on the short runway. When the proportion of heavy planes was low, a simple heuristic was to land most of the planes on the long runway, only using the short runway if necessary due to overload. This mostly avoided any complex calculation about plane size and wind direction.

However if the proportion of heavy aircraft (747s) increased this strategy would break down, since the long runway would now need to be used effectively for the 747s. Schunn and Reder (2001) found that participants were sensitive to their changing success rates and altered their strategy accordingly. They also further discovered that such adaptivity correlated with certain individual cognitive differences as measured with psychometric tests of reasoning. Strategy choice was also affected by different success rates in an earlier study by Reder (1987). In the study two groups were given comprehension questions that could be answered either by a plausibility judgement strategy or a retrieval strategy. Participants in one group given a mix of 80% plausibility and 20% retrieval questions, and subsequently were more likely to use plausibility judgments on later questions. Whereas, participants given a 20% plausibility 80% retrieval mix did the opposite and were more likely to use a retrieval strategy.



This demonstrates that participants were sensitive to what had worked in the immediate past and continued with the successful method. The basic result that participants respond to differing features in the experiment has also been found in other contexts, such as problem solving (Lovett and Anderson, 1996). Thus, it does seem that people are sensitive to the success rate of a strategy and modify their approach accordingly. It is possible then, that IR may be adopted when it is a sub-optimal strategy, though its use may depend on its overall level of success within the task.

### 7.1.1 RATIONALE

IR is proposed as a general part of skilled performance in some tasks that have an element of intrinsic redundancy. If such reduction strategies are part of real world skills some stimulus variability should be tolerated. One such variable is the regularity of the redundancy; that is does the redundant information have to always be 100% redundant at all times. The evidence above suggests that people do use strategies that are sub-optimal and also rely on information sources that are not 100% reliable. It also seems that individuals are sensitive to the success rate of their performance and can select strategies appropriately.

Whether these properties are true of IR is an empirical question yet to be tested. It may be that IR only develops when the redundancy is entirely regular across trials. Conversely, it may be that a large degree of irregularity is accepted over source reliability. To put it another way, IR is an operand that may or may not be employed dependant on its success rate over the immediate past. Success rate of the IR strategy can be manipulated during training by varying the number of alphabetic errors that occur post-triplet. If none occur in this way, then the strategy will be 100% effective for correct and incorrect stimuli. In contrast, as more alphabetic errors occur outside the triplet, if this is all that is processed, the success of the strategy will achieve a corresponding decline.

#### *Levels of regularity*

Deciding on the irregularity of the redundancy – how often post-triplet errors occur, is difficult if a limited number of conditions are to be realised. A 100% regular condition is necessary for a comparison. At the other extreme, in all the experiments described that incorporated the usual test phase, IR was reversed when post-triplet errors were re-

introduced at a 60% level. This suggests that participants start to process the whole string again when 30 out of the 50 incorrect strings had post-triplet errors. This is not very surprising as feedback was after each trial indicating an often failing strategy to the participant. A 60% level is still interesting though, since participants may still reduce if there is not a sudden change, but rather this level is held constant over all blocks/trials.

A further level of regularity between these two extremes is suggested by the research reviewed earlier. A probability switch around the 80% level seems to promote if not boundary changes, then some criterion change (Ratcliff, 1981). Similarly both task performance and trust following automated aid seems to fail around this 80% level of reliability. Reder (1987) again used an 80% - 20% split and found people sensitive to the change. Therefore, there is some, albeit slight, indication that a change in strategy occurs around this point. Since the aim in running the experiment was to test if IR strategies tolerate some irregularity in information source redundancy, the intermediate level selected was 90% regularity. At this proportion, if IR can occur with any irregularity in redundancy, it should in this case. Given unlimited participants, additional levels would have been worthwhile, since this was not the case, it was believed that 100%, 90% and 60% gave an interesting comparison for the reasons explained.

### *Feedback conditions*

In the AAT and TST paradigms described it was common for participants to receive feedback after each and every trial. While it is common for many real tasks to be like this, it is also true that often performance feedback will not be immediate, but rather some time 'downstream'. For example, many mechanical failures due to operator error do not occur immediately after servicing but some time later. The timing of error feedback is thus an interesting variable as far as IR is concerned. Haider and Frensch

(1999a) have utilised end-of-block feedback in one set of their studies. The reason they report for doing this was that they were particularly interested in error rates, and this provided the best sensitivity for this measure. During their task they found that IR progressed as usual with this change in feedback timing. So it seems that it is not necessary for individuals to have feedback on their own performance levels after each trial for IR to occur.

Since the change in feedback timing does not appear to discourage IR in the usual set-up, then it was thought advantageous to also adopt feedback at the end of each block for this study. The main reason for this is that if participants started to get error feedback as soon as they tentatively began to reduce, it was unlikely that they would adopt an IR strategy further into the session. Certainly, for the 60% condition receiving error feedback following each trial at a high level similar to the usual test phases was likely to stop IR as it developed. End-of-block feedback when used should therefore allow even slight shifts towards IR to be detected. End-of-block feedback is also likely increase sensitivity when measuring IR in the test phase. If participants do ignore post-triplet segments at all, then they will miss and continue to miss post-triplet alphabetic errors for the whole test block under these conditions.

One of the reasons for running the first study in the thesis was to determine whether IR was adopted for more ‘noisy’ stimuli. For the first experiment, letter case was randomised so that each sting differed visually across trials even though the actual letter strings did repeat in a block. For example;

E [4] j K

e [4] J k L

Under these conditions IR was still found. So, it seems that extremely consistent experimental stimuli were not necessary for the effect, or individuals can tolerate some changes provided those changes are superficial. The final study reported in this chapter is a more extreme test of this rationale; the constancy of source redundancy was manipulated so that post-triplet errors were not always task-irrelevant. In addition the timing of feedback was also changed, so it occurred only at the end of each block, not after each trial. While Haider and Frensch (1999a) found IR developed under these conditions, their participants worked under very specific task instructions. It was felt the change in feedback could perhaps influence the way participants approached the *standard* task, so a pilot study using feedback at the end of each block was run to determine its effect.

The AAT was selected for the following experiments to make the results directly comparable to published studies demonstrating IR (Haider and Frensch, 1996, 1999a, 1999b). These studies found a strong reliable effect of IR, suggesting the AAT would be ideal to determine if changes in the regularity of error positioning can disrupt IR or not.

## **7.2 EXPERIMENT 5a: PILOT STUDY: FEEDBACK AT THE END OF A BLOCK**

As indicated, changing the paradigm so that participants received feedback at the end of a block, instead of after each trial, changes the task in a number of ways. First, participants do not know whether they are performing correctly through the block, which may lower confidence in reducing. Second, when feedback was received at the end of a block, it would not be possible for participants to determine which strings the errors came from. Again, this could effect the development of an IR strategy. Since

these changes may fundamentally affect IR it was decided to run a pilot study in which just the feedback change was made. 14 participants were run, employees of the Open University, aged between 25 and 60 with a mean age of 41.35. None had done either the AAT or TST before. The design and procedure closely followed that of Experiment 1 in the first thesis study, except of course feedback now occurred at the end of each block.

### **7.2.1 PILOT RESULTS**

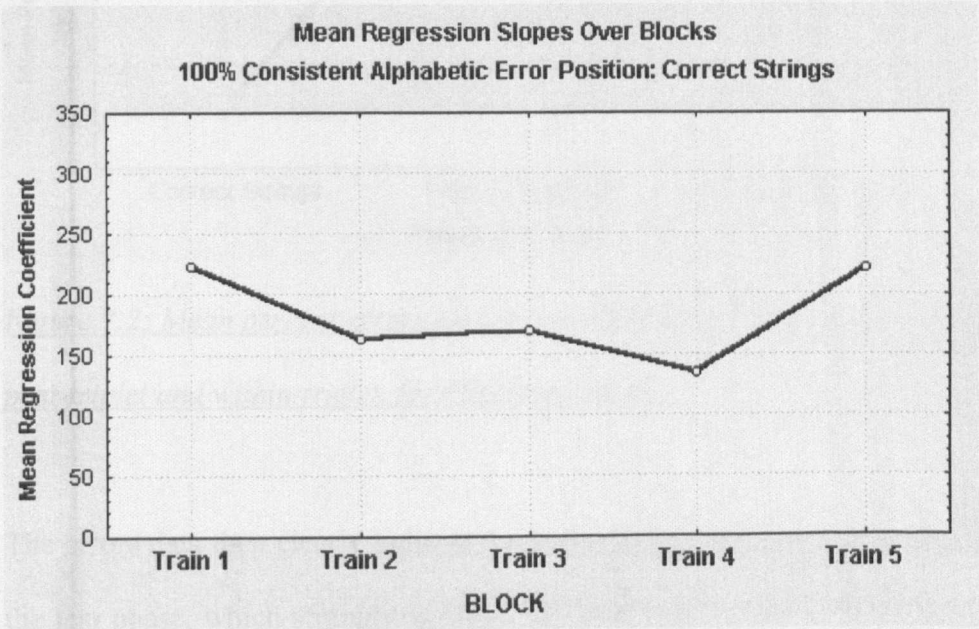
As before the areas of interest were how regression coefficients declined over training blocks and also how error rates were affected when alphabetic errors occurred post-triplet in the final test phase. To test this, once again for the pilot, regression coefficients and percentage errors for the test block were calculated.

#### **REGRESSION COEFFICIENTS: CORRECT STRINGS**

Figure 7.1 shows mean coefficients over training blocks. The change over blocks was not as expected and did not appear to reduce across blocks. The data were subjected to a One-Way ANOVA with Training Block 1-5 within subjects as the independent variable and mean coefficient as the dependant variable. No main effect of Training block was found  $F(4,52) = 1.88$ ,  $MS_{\text{error}} = 11037.02$ ,  $p = 0.127$

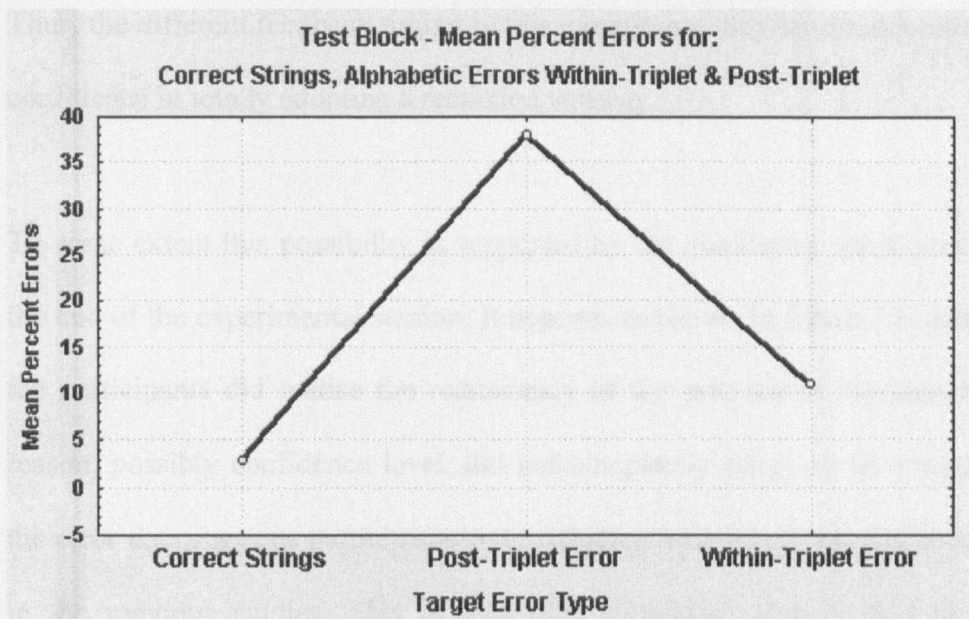
Further analysis compared first to last test block, since this was where a reduction in coefficient was particularly expected. One-Way ANOVA with Training Block first –last as the independent variable, again found no main effect of training block.  $F(1,13) = 0.00066$ ,  $MS_{\text{error}} = 19642.12$ ,  $p = 0.9799$ . Similarly comparing the last training block to the final test block coefficients found no effect Block(last Training vs. Test – within subjects)  $F(1,13) = 0.225$ ,  $MS_{\text{error}} = 33069.31$ ,  $p = 0.643$

These results indicate regressions slopes did not decline over training blocks and also there was no rise in slopes when the rule changed during the test phase. This directly indicates that there was no reliable reduction in the string-length effect over training. Indirectly, this also suggests that IR did not occur during the training phase.



*Figure 7.1: Mean regression coefficients over training blocks for correct strings where alphabetic error occurred 100% within the triplet, for Pilot Experiment.*

If this is the case then participants should not make more errors during the test block when the alphabetic error occurred outside the trained position; post-triplet. To test this, percentage errors for, trained, novel alphabetic error positions and for correct strings were calculated and are shown in figure 7.2. One-Way ANOVA with Position as the independent variable Position(Trained, Novel & Correct Strings – within subjects), found a main effect of Position  $F(2,26) = 6.516$ ,  $p = 0.0051$ . Post hoc (Tukey) finds that percentage errors when the alphabetic error occurred post-triplet was reliably higher than for within-triplet errors and correct strings,  $p = 0.036$  and  $p = 0.005$  respectively.



*Figure 7.2: Mean percent errors during test phase for correct strings, alphabetic errors post-triplet and within-triplet, for Pilot Experiment.*

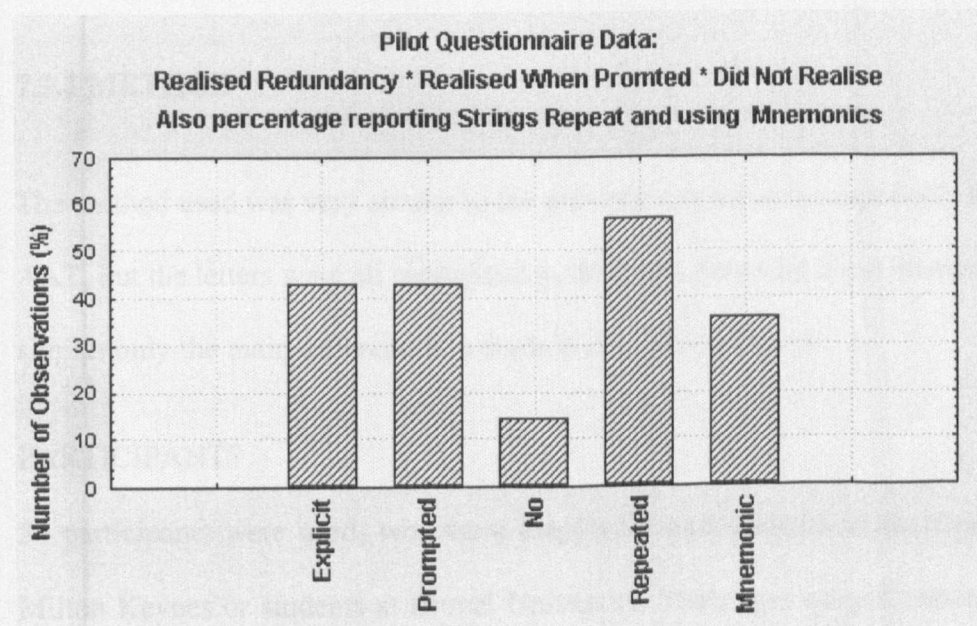
The errors data then clearly indicate participants were missing post-triplet errors during the test phase, which strongly suggests that they had reduced and were not processing the post-triplet segment consistently. Thus, the regression slope and error data contradict one another. However, the regression slopes are a less direct method of measuring IR. If participants checked the post-triplet segment often enough, slopes may decline less over blocks. That is, to some extent string-length would predict RT to a greater degree. This may make the change in string-length effect over blocks more difficult to find.

Accepting this, participants could still however miss many of the post-triplet test errors, as they were not checking every string completely every time. This is a logical conclusion that also explains why changing feedback timing caused this change. That is, since feedback only came at the end of a block, participants were not sure of their



performance and so may have intermittently checked the post-triplet part of the string. Thus, the different feedback timing in this experiment may have moderated participants confidence in totally adopting a reduction strategy.

To some extent this possibility is supported by the qualitative questionnaire data from the end of the experimental session. It appears, as shown in figure 7.3, that a number of the participants did realise the redundancy of the post-triplet segment, but for some reason, possibly confidence level, did not completely adopt an IR strategy. However, the error data suggests participants were reducing ‘at a level’, but not as consistently as in the previous studies. This is somewhat surprising, since within the Haider and Frensch study (1999a), IR did occur under end-of-block feedback conditions, especially under speeded instructions.



*Figure 7.3: Responses to questionnaire data from Pilot Experiment:*

- i. Those who explicitly realised post-triplet redundancy,
- ii. Those who realised it when prompted and
- iii. Those who that did not know of the rule.
- iv. Also shown are those who reported using mnemonics and thought the string repeated.

One obvious way to counteract the drop in effect, while retaining end of block feedback, is to put participants under some extra pressure to adopt IR. Since this type of feedback was used successfully in the Haider and Frensch (1999a) study, speed pressure presents as one alternative. So, in the main study participants were told to work through the strings as fast as possible. The hypothesis were

- i. IR should occur in the 100% condition under speeded instructions.
- ii. If IR operates as a heuristic, reduction should occur when error position regularity was reduced to 90%.
- iii. Reducing error position regularity to 60% was expected to discourage the development of the reduction strategy..

### **7.3 EXPERIMENT 5b: MAIN STUDY: MANIPULATING THE REGULARITY OF SOURCE REDUNDANCY**

#### **7.3.1 METHOD**

The method used was very similar to the previous experiments, especially Study 1 – AAT, but the letters were all capitalised in this case. Since the experiments were so similar only the main differences in method are reported below.

#### **PARTICIPANTS**

54 participants were used; who were employees and students of the Open University Milton Keynes or students at Brunel University. Their ages ranged between 18 and 59 with a mean age of 34.32. For participation each was paid £10. No one had attempted either the TST or AAT before

## STIMULI AND PROCEDURE

The strings were generated as for Study 1, except that for the *training* blocks, of which there were five, the position of the alphabetic error was not always within the triplet e.g.

100% Condition 1. All errors are within the triplet.

90% Condition 2. 5 of the 50 incorrect trials have errors outside the triplet.

60% Condition 3. 20 of the 50 incorrect trials have errors outside the triplet.

When the error occurred post-triplet during the training blocks, the position of the error was balanced in position as far as possible throughout this segment of the strings.

For the following test phase as in Study 1, 20 of the 50 incorrect test trials had alphabetic errors post-triplet. This made this block the same as the 60% Condition blocks and made all the 60% Condition training and test blocks the same in this respect.

Feedback for all blocks came at the end of a block in all conditions, to overcome the problems this caused in the pilot study, participants were instructed to work through the strings as quickly as possible. This instruction was given by the experimenter during the initial briefing and also in written form at the start of each block.

At the end of the session in addition to the usual final questionnaire, participants were asked to judge how much they thought the task was to do with speed vs. accuracy. They did this by marking a line to dissect a 10cm Likert scale anchored with the words 'speed was important in this task' at the left end of the scale and 'accuracy was important in this task' at the other end. The purpose of this was to determine how the task was approached especially in light of the 'work quickly' instructions.

### 7.3.2 RESULTS

#### OVERALL ANALYSIS

A slightly different approach was taken to excluding outliers from the analysis in this experiment. Since this study directly investigated the conditions under which reduction will or will not occur, it is particularly important to exclude participants who did not undertake the task properly. Previous experience with the AAT suggested that sometimes people focus on the triplet as the whole task, never really processing the post-triplet segment at all. To remove this confounding variable, participants that start out with a mean regression coefficient of less than 15 were removed from the analysis. Such a low coefficient indicated string-length was not affecting reaction time. One could argue these participants were not processing the whole string even at the start of the experiment.

As before data from participants making excessive errors per block were also removed. In this latter case the criterion was set at a 15% error rate. This was higher than previous experiments, but was to allow for missing post-triplet errors in the 90% condition. If participants did start to reduce in this condition they would miss post-triplet string errors, but not through poor performance. Therefore the criterion was raised by 5%. This was a similar level used by Haider, Frensch and Joram (In Press) in their comparable study. This resulted in  $n = 14$  for Condition 1,  $n = 15$  for Condition 2 and  $n = 15$  for Condition 3 (Participants removed from Cond 3., all had very low first block slopes ( $n=2$ ) or error rates over 40 per block ( $n=1$ )).

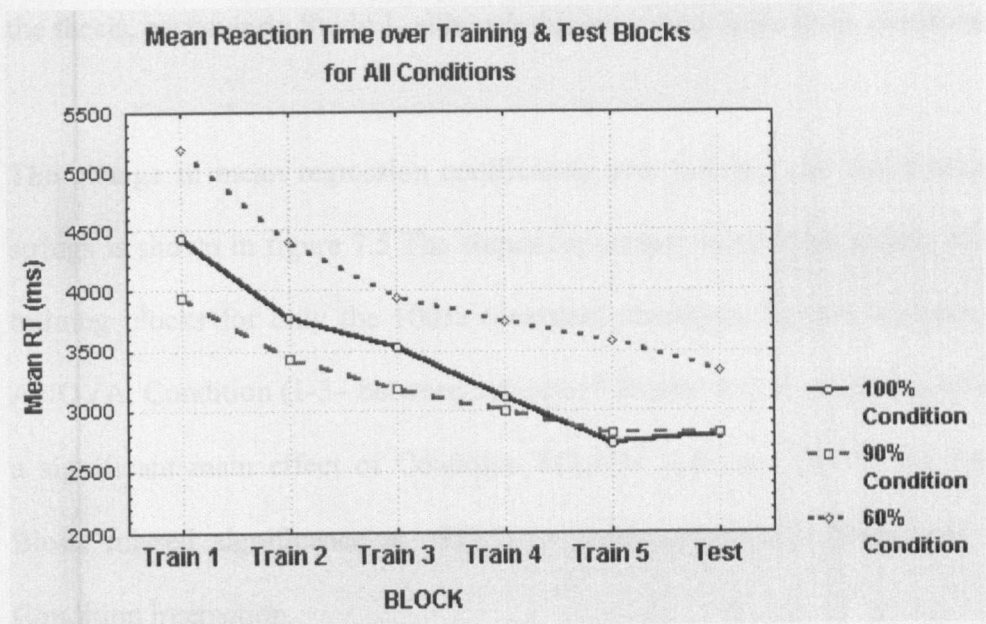
There was no speed-accuracy trade-off. The correlation between RT (Reaction Time) and mean percent error rate for the 3 conditions were (excluding responses to strings where the error could occur in the redundant segment);

- (i) Condition 1(100%),  $r(14) = .374$ ,  $p > 0.05$
- (ii) Condition 2 (90%),  $r(15) = .357$ ,  $p > 0.05$  and
- (iii) Condition 3 (60%),  $r(15) = -.101$ ,  $p > 0.05$

Figure 7.4 illustrates the mean RT for responses to all strings for each condition. RT appears to reduce over training blocks for all three conditions. 2-Way ANOVA Block (1-5, within subjects) \* Condition(1-3, between subjects) revealed a main effect of Block  $F(2,41) = 6.52$ ,  $p = 0.003$  and a main effect of Condition  $F(4,164) = 93.51$ ,  $p < 0.0001$ . The Block \* Condition interaction just missed significance at  $F(8,164) = 1.95$ ,  $p = 0.0563$ .

Post-hoc tests (Tukey, for unequal N) revealed a significant reduction in mean RT between the first and last training blocks for each condition ( $p < 0.05$ ). The 60% condition also had significantly higher mean RT at the first and last training blocks than either of the other two conditions ( $p < 0.05$ ), while for these blocks the 100% and 90% conditions did not differ reliably.

Further analysis of mean RT for the conditions between the last training and final test block found no reliable increase for any group. To some extent this was to be expected as feedback was at the end of a block, so participants in the 100% and 90% conditions would not receive feedback immediately they made a mistake so would not be alerted to the test block change. For the 60% condition all blocks, including test, were equivalent, so again RT was not expected to change for the final block.



*Figure 7.4: Mean Response Time (RT) for string verification over all blocks for the three conditions of Experiment 5b.*

The pattern of results suggests that participants became faster at verifying strings as training progressed. Participants in the 60% condition were overall slower; the parsimonious explanation for this is that they were checking the whole string more thoroughly as the post-triplet segment was more relevant than in the other two conditions. Or to put it another way, the rule was unreliable and so they could not safely ignore the post-triplet items.

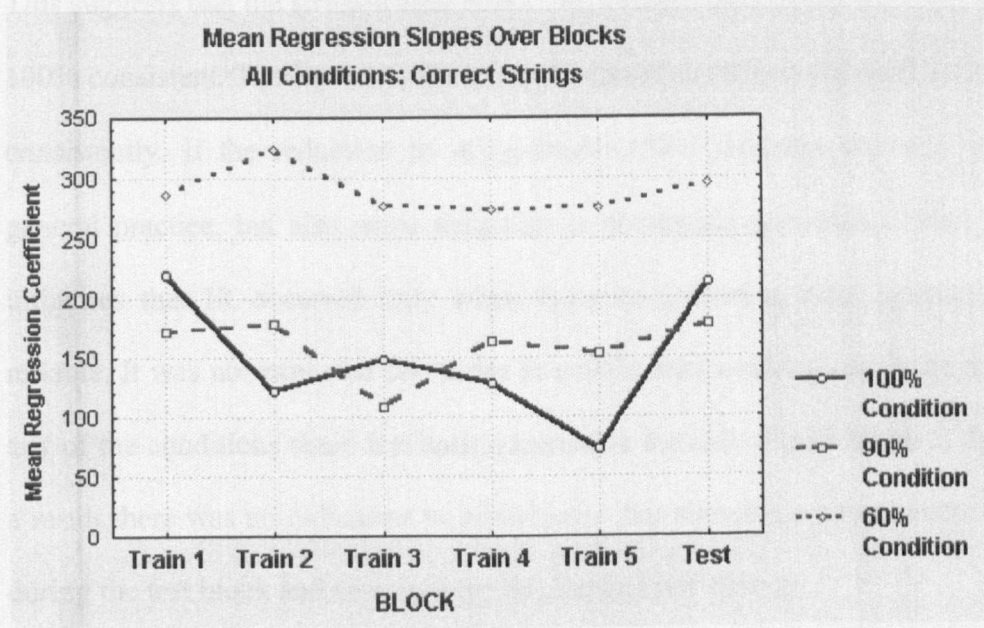
## INFORMATION REDUCTION WITHIN THE CONDITIONS: REGRESSION SLOPES CORRECT STRINGS

As before, for all conditions the best fitting linear regression slopes were computed separately for each subject and each trial block using string length as the predictor and reaction time as the dependant. The mean regression coefficients obtained were then



subjected to ANOVA. The analyses followed the same form as the previous studies in the thesis, particularly Study 1, although this time there were three conditions.

The change in mean regression coefficients over training and test blocks for correct strings is shown in figure 7.5 The slopes for correct string trials appear to decline over training blocks for only the 100% consistent condition. Further analysis with 2-Way ANOVA, Condition (1-3– between subjects) \* Block (1 to 6 – within subjects) revealed a significant main effect of Condition  $F(2,41)= 9.20$ ,  $p= 0.0005$ , the main effect of Block missed significance at  $F(5,205)= 2.05$ ,  $p= 0.0732$ . There was no Block \* Condition interaction.



*Figure 7.5: Mean regression slopes over Training and Test blocks for all conditions: Correct Strings, Experiment 5b.*

Additional analysis within each group with training block as the independent variable comparing first to last training blocks only: 1-Way ANOVA Block(First Train, Last

Train – within subjects), found a significant reduction in slope only for the 100% consistent condition, see Table 7.1 below.

maintained in the discussion.

*Table 7.1: Analysis of First vs. Last Training Block within each condition, Experiment 5b.*

Condition	Main effect of Block first vs. last training block	
100% Consistent	F(1,13)= 19.45	p = 0.0007
90% Consistent	F(1,14)= 0.488	p = 0.496
60% Consistent	F(1,14)= 0.035	p = 0.855

One-Way ANOVA's

This suggests that the string-length effect only reliably declined when error position was 100% consistent. That is, when the errors for incorrect strings occurred within the triplet consistently. If the reduction in string-length effect indicates not only the effect of general practice, but also some reduction in processing post-triplet items, this further intimates that IR occurred only when the rule governing error position was 100% reliable. It was not expected that a rise in coefficients would occur at the test phase for any of the conditions since feedback occurred at the end of each block in this study. As a result there was no indication to participants that alphabetic errors were being missed during the test block and so no prompt to change their strategy.

coefficients are smaller than for the 100% group.

However the slopes in Figure 7.5 do not support this, the 100% condition seems to rise markedly during the test phase. To determine if the change was reliable, within each condition the last training block was compared to the test block: One-Way ANOVA's Block(Last Train, Test – within subjects) was again run. The results in table 7.2 below indicate that the rise in regression coefficients was significant only for the 100% group.



This rise is interesting as it appears participants did notice a change in error position regularity during test even without immediate feedback. This point will be further examined in the discussion.

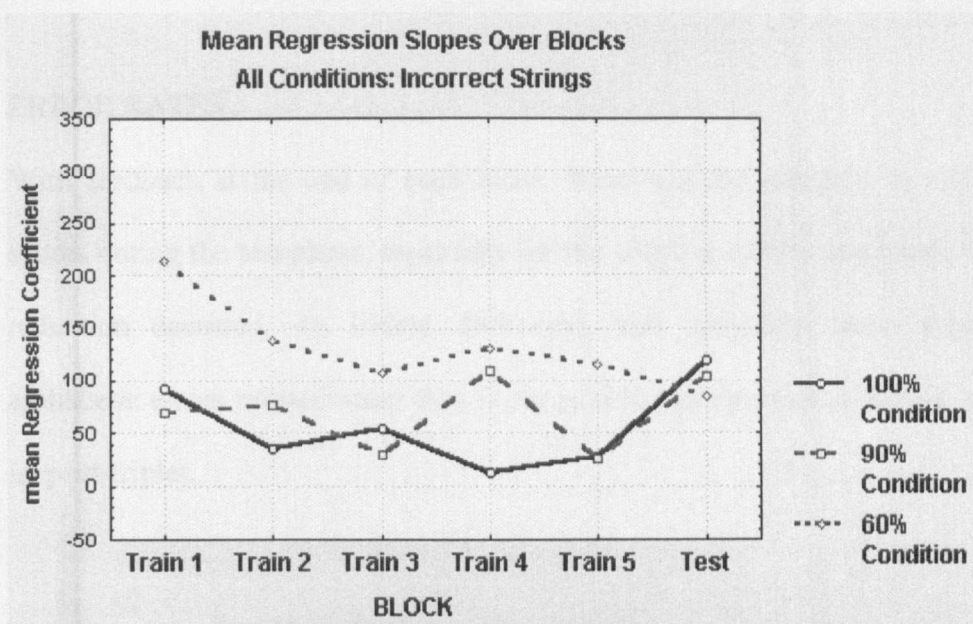
*Table 7.2: Analysis of Last Training vs. Test Block within each condition, Experiment 5b.*

Condition	Main effect of Block last training vs. test block	
100% Consistent	F(1,13)= 12.20	p = 0.004
90% Consistent	F(1,14)= 1.016	p = 0.331
60% Consistent	F(1,14)= 0.165	p = 0.690

Overall the pattern of results indicates the only reliable change in coefficients over training and test occurred for the 100% condition. Subsequently, this should also lead to more errors for this condition when alphabetic errors occur outside of the usual position during the test block.

### INFORMATION REDUCTION WITHIN THE CONDITIONS: REGRESSION SLOPES INCORRECT STRINGS

The mean regression slopes for incorrect strings are shown in figure 7.6 overall the coefficients are smaller than for the correct strings as expected. The data were subjected to a 2-Way ANOVA, Condition(1-3– between subjects) \* Block(1 to 6 – within subjects) and revealed a significant main effect of Condition F(2,41)= 3.347, p= 0.045, there was no main effect of Block or Block\* Condition interaction.



*Figure 7.6: Mean regression slopes over Training and Test blocks for all conditions: Incorrect Strings, Experiment 5b.*

Additional analysis within each condition with training block as the independent variable comparing first to last training blocks only: 1-Way ANOVA Block(First Train, Last Train – within subjects) found no significant effect of block for any of the conditions. Similarly, analyses comparing last training block to test block mean coefficients found no reliable difference.

However, the 90% condition just misses significance at  $F(1,14)= 4.58$ ,  $p = 0.0504$ . The rise in coefficients for the test block was also found in previous studies in the thesis. One possible explanation is that when the rule changes during test and errors are made, people become confused and start to check the post-triplet string for both correct and incorrect strings. Again this was surprising in this study as feedback was not given until the end of the test block, so participants should not have been alerted to the change. Unless of course they tended to check the post-triplet string intermittently all through the study.

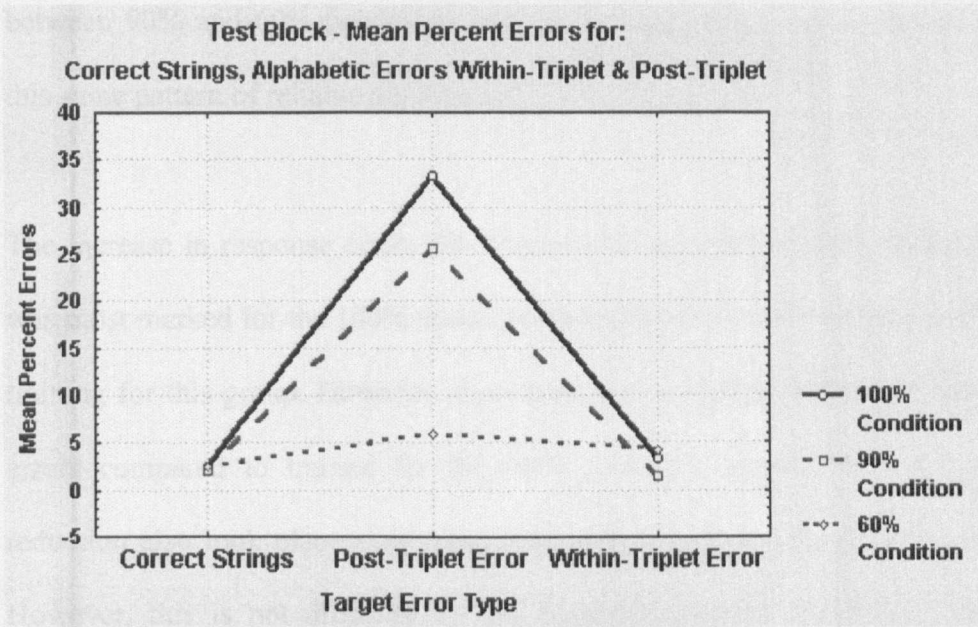


was a reliable main effect of Position(Trained, Novel & Correct strings)  $F(2, 10) = 10.56, p < .001, \eta^2_p = .68$ .

ERROR RATES

With feedback at the end of each block, there was the potential to make high error scores during the test phase, especially for the 100% and 90% conditions if any level of reduction occurred. As before, following such reduction there should be more alphabetic errors missed when they occur in novel as opposed to trained positions, that is, post-triplet.

mainly higher than the 60% condition, especially for the novel position.



EXPERIMENT 5b: EFFECT OF FEEDBACK ON THE REDUCTION OF ALPHABETIC ERRORS

*Figure 7.7: Mean percent error rate for the final test blocks, target in the trained or a novel position; correct string response errors added for comparison, Experiment 5b*

There does appear in figure 7.7 to be a difference in percentage errors between conditions when the target appeared in a novel position. 2-Way ANOVA Condition(100%, 90% & 60% – between subjects) \* Position(Trained, Novel & Correct strings – within subjects) did not find a significant main effect of Condition, but there

was a reliable main effect of Position  $F(2,82)= 16.689$ ,  $p< 0.00001$ . Also there was a significant Condition \* Position interaction  $F(4,82)= 3.328$ ,  $p= 0.014$

Post-hoc tests (unequal N Tukey) found reliably higher response errors for novel than trained position alphabetic errors for both the 100% ( $p<0.001$ ) and surprisingly the 90% condition ( $p<0.01$ ). There was no significant difference between these error positions for the 60% condition. Regarding just the novel position errors, the 100% condition was reliably higher than the 60% condition ( $p< 0.01$ ), but there was no reliable difference between 90% and 60% conditions ( $p= 0.07$ ). Less strict paired comparisons still find this same pattern of reliable differences.

The increase in response errors for strings were alphabetic errors occurred post-triplet was most marked for the 100% condition as expected from the drop in coefficients over training for this group. However, there was also a reliable increase in ‘novel position’ errors compared to trained for the 90% condition, suggesting that some level of reduction also took place even when the underlying rule was not completely regular. However, this is not mirrored by the regression results so is less unequivocal. As expected, there was no such rise in percentage error rate for the 60% condition, which included post-triplet errors at the same percentage level throughout all the training and test blocks.

## END OF SESSION QUESTIONNAIRES

As with the studies in previous chapters, participants were asked set questions to determine if they noticed anything about the strings and were prompted towards the end to determine if they had noticed anything about the position of the errors (see Appendix

2). One participant did not want to take part in this final stage of the experiment due to the length of the session. Figure 7.8 illustrates participant's responses.

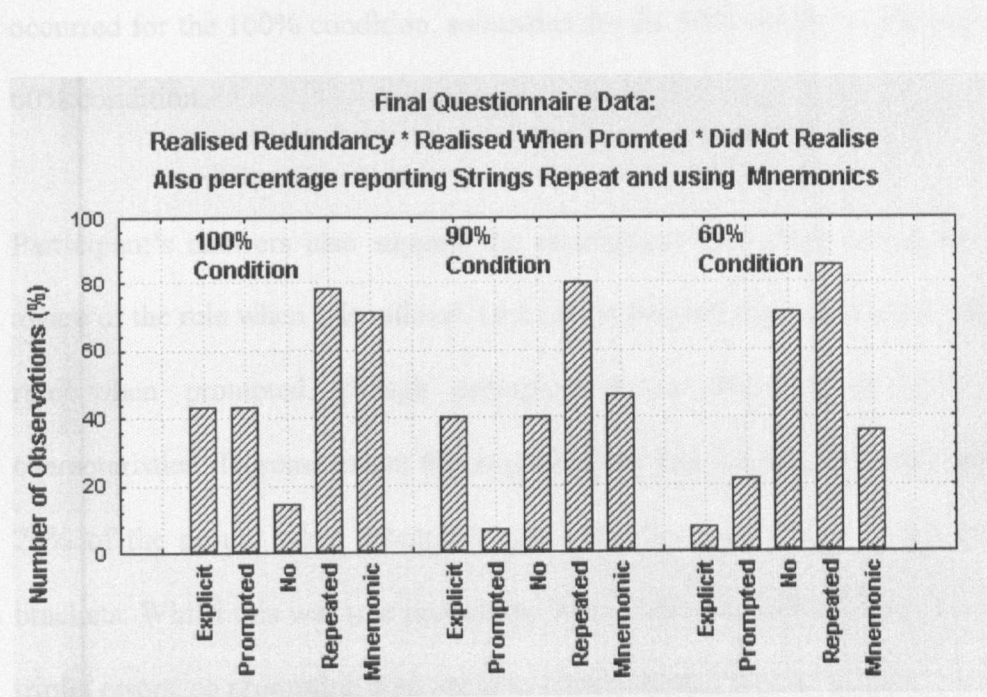


Figure 7.8: Participant responses to final questionnaire, Experiment 5b;

- (i) Those who explicitly realised post-triplet redundancy,
- (ii) Those who realised it when prompted and
- (iii) Those who that did not know of the rule.
- (iv) Also shown is the number that reported using mnemonics and thought the strings repeated.

The qualitative data support the earlier statistical analysis. More participants realised the underlying rule determining error position when the rule was 100% consistent than for the other two conditions, especially when the prompted reports are included. A number in the 90% condition still reported they thought the alphabetic errors followed the brackets, which of course was mostly true. In contrast very few noted this for the 60% condition, which again was correct; errors were almost as likely post-triplet as within

the triplet. These reports mostly suggest the experimental manipulations were noticed by participants, and support the regression and error analysis, which indicated IR occurred for the 100% condition, somewhat for the 90% condition and not at all for the 60% condition.

Participant's answers also support the assumption that often people are consciously aware of the rule when it is utilised. One caveat here though is that some only realise the rule when prompted, though prompted report obviously is open to demand characteristics. To some extent this is particularly true for the 60% condition; more than 20% of the group, when prompted, agree that the error tended to always follow the brackets. Whilst this was true percentage wise, there was a 60/40% split for triplet/post-triplet errors, so prompting does seem to encourage this type of report.

As noted in thesis study 1, participants do seem aware that the strings repeat. Also as before, a number of participants said they developed mnemonics to aid performance. The use of these mnemonics appears to decline as rule regularity falls. This seems logical as such mnemonics are less likely to always result in the correct response when rule regularity falls to 90% and 60%.

#### SPEED/ACCURACY AND TIME PER SESSION

Participants were also asked to indicate by use of a Likert scale (see Appendix 2) whether they thought the task demanded mostly speed or accuracy. This was to determine the effect of the speeded instruction given during the experiment, but also gives some indication of how participants approach the task. Statistical analyses of the scale responses found no significant difference in scores between groups. Measured on a 10cms scale the mean percentage scores were 58.2%, 64.5% and 64.2% for 100%,

90% & 60% conditions respectively. The scores were biased towards the accuracy end of the scale.

Each session was also timed from the start of the first training block to the completion of the test block. One-way ANOVA found a reliable difference between groups,  $F(2,40) = 17.37$ ,  $p < 0.0001$ . Post-hoc Tukey for unequal N, indicated that significantly longer was taken to complete the blocks only for the 60% group  $p < 0.001$ . the mean times were 51.01, 47.22 and 65.5 minutes for 100%, 90% & 60% conditions respectively. The most likely explanation for this is that trials took longer in the 60% condition since all the string had to be verified constantly.

### **7.3.3 DISCUSSION**

#### *Main results*

- (i) The Speed manipulation successfully produced a reduction in slope over blocks for the 100% condition, so removed problem found in the pilot.
- (ii) The string-length effect only reliably declined when the error position was 100% consistent.
- (iii) There was a Rise in coefficients for 100% condition at the test block even without immediate feedback.
- (iv) Increase for novel position errors for 100% & 90% conditions, somewhat in contradiction to the regression analysis.
- (v) People seem consciously aware of the rule and are sensitive to the experimental manipulation.

The purpose of running the final study of the thesis was to determine whether IR operated as a sub-optimal strategy. Considering the research that found sub-optimal methods, such as heuristics, are often employed, this seemed possible. The hypothesis was then, that IR would certainly develop when the error position was entirely regular. However, if IR could operate as a sub-optimal strategy, reduction should also occur when error position became irregular 10% of the time. At levels of irregularity higher than this, it was thought IR would not develop since it would not work for, in this case, 40% of the time for incorrect strings.

One of the main overall concerns with the research investigating IR, was the generalisation of experimental findings to actual skilled performance. The paradigms used, such as the AAT and TST, were not only much reduced versions of any real task, they were also entirely consistent in the way relevant and irrelevant information was determined. This consistency is unlikely in a 'real world' setting and stimuli may vary on any number of dimensions, including item relevance. So it was interesting to determine how regular task redundancy needed to be for an information reduction strategy to develop. Thus, the aim was to find IR under sub-optimal conditions, if indeed reduction did then occur to any extent. To this end the task was designed to be as sensitive as possible to IR by providing feedback only at the end of a block. It was thought that if there were any potential for IR it would be found and measured using this method.

#### *End-of-block feedback and residual processing*

Regarding the pilot study, which had 100% regular strings, end-of-block feedback had some unexpected results. While test error scores suggested IR had occurred, the



regression slopes did not flatten over practice as expected. The latter finding does not indicate any reduction in processing over practice. This prompted the use of speeded instructions to further encourage IR. This it appears, had the desired effect and both the regression and error analysis seemed to demonstrate IR developing in the main study for 100% regular strings.

In itself, this was an interesting result; IR did develop when feedback was not immediate, but seemed to require some additional pressure before it was fully adopted. Moving the feedback to the end of a block rather than after each trial most probably affected the confidence in using the strategy fully. This is understandable when the change is considered, not only did participants not know their accuracy from trial to trial, but when they did get the feedback, it did not inform which strings they had made an error with. Under these conditions developing any strategy is comparatively a riskier endeavour. One could propose that this led to ‘intermittent’ checking of the whole string, or to use a previous term, there was a high level of ‘residual’ processing. This in turn meant that string-length remained a predictor of response times over blocks.

This is not to say there was no evidence of IR in the pilot, there was, since error rates in the test phase were higher for alphabetic errors in the novel (post-triplet), compared to trained (within-triplet) positions. However, residual processing also accounts for such error rates. The argument is that participants were intermittently checked the post-triplet segment, sustaining a string length effect. This checking was not frequent enough though, to catch all of the post-triplet errors in the test block.

This argument is more acceptable if the residual processing results from the previous experiments in chapters 4-6 are examined. The main effect of the manipulations across

these studies seems to be moderation of the 'level' of IR. For example, IR did not really develop much earlier when difficulty or familiarity was changed, rather target absent string slopes just became closer to those found with target present strings (which as previously inferred may be regarded as a baseline). The real test of this would require many more blocks, to determine if all conditions would eventually reach the base line. However, buying into the proposal that various manipulations do not so much moderate when IR develops but rather to what extent it is adopted, then the pilot results for regression and error analysis are understandable. Checking of the irrelevant string segment was attenuated, but to a somewhat lesser degree than expected during the pilot study.

#### *Levels of error-position regularity.*

Under instructions to work quickly in the main experiment, regression slopes declined over blocks as expected when string regularity was 100% consistent. At this level of consistency, participants also made significantly more test errors when the alphabetic error occurred in a novel post-triplet position. This replicated earlier findings by Haider & Frensch, (1996, 1999a, 1999b), and thesis Study 1. So, IR appears to develop for this condition. In contrast, when the position of relevant string items was less regular, at 60%, neither did the regression slopes decline over blocks or was there a rise at test for post-triplet target misses. It can be concluded from this that IR did not occur for this condition. Thus, at each extreme of the levels of regularity tested, the results were clear. When entirely regular strings were presented, IR developed, but this was not the case at much lower levels of regularity.

The 90% condition was however equivocal; regression slopes did not decline over blocks, but there was an increase in test errors for post-triplet targets. When 10% of the

incorrect training strings contained alphabetic errors post-triplet, it seems there remained a string-length effect. So, participants were processing the post-triplet string segment at least to some extent. However, in the final test block when 40% of the incorrect strings held post-triplet alphabetic errors, enough were missed to increase error rate for this irregular string type. Specifically, there were significantly more errors for post-triplet targets than for within-triplet targets during the test phase.

This was a similar result to pilot study, and begs the same argument. The parsimonious explanation for the incongruent regression and error analysis is possibly that participants were intermittently scanning the redundant string segment. They did this enough to sustain a string-length effect, but not enough to catch all the test targets. Compare this to the 60% condition, which showed neither regression decline over blocks or increased error for the irregular test strings. This rather indicates some IR was taking place in the 90% condition, but at an attenuated level.

#### *Test block errors and residual processing*

Further support for this view was provided by the test-block performance in the 100% condition. In this condition the regression coefficients declined over blocks as indicated, but increased again reliably for the test block. In previous studies this was expected, and results from participants realising the strategy was no longer valid. The change in regularity at test become apparent (if they had reduced), since error-feedback *after each trial* informed participants that they were missing post-triplet targets. Consequently, they started to check the whole string. The current experiment was different though, since error feedback only occurred at the end of the test block. So there would be no prompt that the strategy was failing. The only way this could be detected was if some residual checking of the post-triplet segment took place at some attenuated level. This

views IR as a strategy that is adopted to different levels rather than an 'either or situation' where one reduces or not.

### *IR as a sub-optimal strategy*

Accepting this, it seems that IR did occur as a sub-optimal strategy in the 90% condition. It could be argued that error rates were a more direct indication of IR. When post-triplet errors were missed, the most likely reason was that individuals were not attending to this string segment, how else would they miss the alphabetic errors there? The regression slopes in comparison infer the string-length effect, which then may indicate reduction. As suggested, both intermittent checking of irrelevant items and practice effects may play a role here in addition to IR. Thus, the error data may be the most reliable measure of IR. So on this argument, the difference between 100% and 90% conditions was really very small.

This agrees with research reporting that people form and use heuristics; rules of thumb which utilise less than all the available information. However, even when just a small number of alphabetic errors were introduced into the mostly irrelevant segment of the strings, the degree to which IR was adopted appeared to drop. Thus, while there was evidence that individuals did reduce, even when the error position was not entirely consistent, they did seem sensitive to this change in regularity. As a result they modified their strategy accordingly. The argument here is that this took the form of increased residual processing of the post-triplet items.

The questionnaire data seem to support this argument, most explicitly reported the 'rule' for the 100% condition, but almost as many did this for the 90% condition and certainly more than those who did in the 60% condition. This gives some suggestion that the rule

had been learned in the 90% condition, even if it was not used as consistently. Overall, it seems that information reduction was developing even if the strategy could not be applied across all the problems. Thus, this study provides further evidence of the conditions under which IR may occur.

*Haider, Frensch & Joram (pre-print)*

Following the completion of this last study in the thesis, during writing Haider, Frensch and Joram submitted a new paper for review. At the time of writing, the paper was not published, but did become available as an electronic proof read article. One of their studies was so similar to the present one here, that it is worthwhile summarising the report.

The thrust of their overall argument was that IR was a conscious choice and so the strategy did not occur as an inevitable automatic result of practice at a task. To put it in their terms, in the case of IR, voluntary processes cause the strategy shifts and not data driven processes. This is somewhat in opposition to contemporary theories of skilled performance. Taking in particular Logan's (1988, 1992; Logan & Etherton, 1994) and also Palmeri's (1997, 1999; Nosofsky & Palmeri, 1997) instance based models of cognitive skill acquisition, the suggestion is otherwise. In these models, strategy shifts are the result of automated learning mechanisms and so come about as an inevitable consequence of practice at a task. Thus, there was a contradiction in the two views.

In their submitted study, Haider, Frensch and Joram (pre-print) conducted a second experiment, which was of most interest, since it was very similar to the main experiment reported in this chapter. The goal of their Experiment 2 was to show that people deliberately chose to use the strategy. One circumstance where they may choose

*not to*, is when the strategy cannot be used for the entire range of problems encountered. To examine this they realised three conditions, always-regular, 90% and 60% regular, where regular referred to how often on incorrect trials the alphabetic error occurred within the triplet part of the string as opposed to post-triplet.

To further determine the voluntary nature of the strategy there was one more manipulation. At the end of each block, participants received 10 correct and 10 incorrect regular trials under highly speeded conditions. Presentation time for these strings was constantly adjusted to the shortest response times occurring for the longest strings in the preceding trials. Thus there was significant speed pressure to adopt a reduction strategy in these trials to save time. Considering the 100% and 90% conditions, Haider, Frensch and Joram (Article pre-print) hypothesised that if IR were a voluntary choice, then one would expect similar performance under the speeded trials if the strategy were available to both groups. However, when this pressure was removed participants in the 100% condition should be more likely to adopt the strategy than for the 90% condition. Participants in the 60% condition were not expected to develop the strategy at all.

This seems to be what they found, regression slopes declined for only the 100% condition, which also showed the highest error rates for post-triplet alphabetic errors during the test phase (However this error rate was also higher in the 90% condition than the 60% condition, so maybe some reduction had occurred?). Under the speeded constraints they found, in contrast to the previous regression analysis, that the 100% and 90% conditions were similar in performance. They concluded that this pattern of results indicated that manipulating the consistency of errors served to affect application of the strategy, but not its acquisition. That is, the strategy was learned for both 100% and 90% conditions, but was only used for the latter condition when forced by lack of time.

*Comparison of Haider, Frensch & Joram (pre-print) and Thesis Experiment 5.*

As can be seen from the description above, there are some striking similarities between their study 2 (Haider, Frensch & Joram, pre-print) and this final study of the thesis. Not only were the levels of irregularity used similar, but also the pattern of results. However the interpretation of these two findings was somewhat different.

Of interest is the 90% conditions in both experiments. Haider, Frensch & Joram (pre-print), suggest that since regression slopes did not reduce over training, the strategy was not adopted for this condition unless forced by lack of time. Experiment 5 in this chapter found a similar result for this condition, which would have led to the same conclusion, if not for the error data. This suggested that more post-triplet than triplet errors had been missed during the test phase. Similarly, Haider, Frensch & Joram (pre-print) found significantly more post-triplet errors were missed in the 90% compared to the 60% condition. It is hard to think of any other reason than IR for this pattern of results. Why else would participants miss so many post-triplet alphabetic errors during the test phase?

A residual processing hypothesis possibly explains the regression analysis in both experiments. To re-iterate, the idea was that under the 90% regular conditions confidence in reducing was lowered, so there was intermittent checking of the irrelevant items. This sustained the string-length effect and so regression coefficients did not reduce over training. This checking was not frequent enough, though, to detect all the increased number of post-triplet errors during the test block. This also then explains the increase in these errors found in both experiments. Accepting this possibility there seems to be evidence that IR occurred for 90% string regularity in both studies.

### *Speed pressure*

While it could be argued that the evidence of IR for the 90% condition in both studies results from speed pressure, some care is needed in this interpretation. Firstly, the instructions to work quickly in our study was not speed pressure in the sense used by previously by Haider and Frensch (1999a) and Haider, Frensch and Joram (Article pre-print). In their studies participants were given just enough time to respond to the longer strings. In their latest study (Haider, Frensch and Joram, pre-print), this was actually linked to performance, so participants never had time to slow on a trial. This was very extreme compared to just an instruction to work quickly, even if this instruction was shown to be effective. Secondly, the results in the thesis study and their latest study both show IR in the test error scores. Importantly, in their study this was the case when their participants were *not* under speeded conditions. So under very mild speeded pressure and in their case no speed pressure at all, people seemed to miss irregular errors in the test block. This tends to support the assumption that IR did develop under 90% regularity even when participants were not really forced to use it.

This both supports the conclusion for Experiment 5 of the thesis and weakens their argument that IR was a conscious strategic decision. Though undoubtedly, there was a change in performance between 100% and 90% regularity. There is also other substantiation that IR was not the inevitable result of practice. One could argue that their speed/accuracy manipulation in an earlier study (Haider and Frensch, 1999a) provided just as good, if not better, evidence that the strategy was under conscious control.



## *Conclusion*

This last study in the thesis extends the view that various manipulations may moderate IR. The mechanism by which the moderation occurs is thought to be changes in residual processing. Rather than describing IR as a switch that suddenly occurs or not, it is best viewed as a strategy where any redundancy is checked less and less frequently as some confidence level in the regularity increases. Under this perspective it appears that IR did develop, even when it was a sub-optimal strategy. That is, it was utilised even if it could not be applied to the whole range of problems. From a practical perspective, this suggests IR may take place in any number of real world tasks and increases generality.

In many respects the results of this experiment were similar to Haider, Frensch and Joram's latest study (pre-print), though the rationale was somewhat different. A comparison of the results calls into question the validity of the 'voluntary process' argument in their case, though there seems to be other sources that suggest IR is most likely a voluntary strategic decision. Of particular interest in the thesis study was the return to processing the whole string during the test block. This occurred even though it was not prompted by error feedback. These points and the residual processing hypothesis will be investigated further in the general discussion to follow.

## **CHAPTER 8:**

### **GENERAL DISCUSSION**

*This chapter brings together the findings from the four main studies, relating the main results to a further understanding of information reduction and contemporary views of skilled performance.*

The thesis set out to examine in more detail some of the quantitative changes in skill acquisition that have recently been reported in the literature. The discussion presented here aims to bring together the experimental findings of each chapter in a meaningful way. The main theme was to illustrate the contribution an information reduction strategy played in an overall picture of skill acquisition and performance. The thesis began with the following old adage, 'practice makes perfect'. As a very general rule of thumb we have learned that practice may not guarantee error-free performance, but it does improve task execution. How it does this has been central to research on skilled performance since the 1940's.

The understanding of skill requires a bridge across both purely theoretical and applied psychology, and across areas that have traditionally been separated such as 'attention', 'learning' and 'skill'. Researchers have tended to concentrate on different parts of cognition. As Logan (2002) reports, a divide-and-conquer approach has prevailed for the last 20-30 years. As a result a number of large and separate literatures have grown. There is a general literature on skill acquisition, but this seems to be separate from research on attention and learning. Strategy selection also seems encapsulated in its own domain.

Logan (2002) has taken a more unified approach, reminiscent of Newell's (1990) SOAR and has tried to integrate certain theories of memory and attention. This is especially interesting, as his instance theory (Logan, 1988) has been one of the most influential accounts of skilled and automatic performance to date. The main tenet of this theory is that over practice a switch is made from an algorithmic solution to a one-step retrieval of the answer from memory. Thus a qualitative change in processing follows practice.

Instance theory stands in contrast to Anderson's (1981) ACT\* production architecture. In case of ACT\*, practice serves to produce ever more refined procedures. As a result, eventually a qualitative change is made from the use of declarative rules to the performance of appropriate procedures directly from memory. The main contrast in the theories is the direction of change between rules and examples as task experience progresses. An instance account describes a change from a rule or algorithmic process to the retrieval of stored examples. The ACT\* architecture, in contrast, describes a change from declarative examples to production rules following practice (Anderson et al., 1997).

Thus, there is a debate here of whether during practice rules lead to examples or examples lead to rules. The thesis may therefore support either theory or some other framework. These two major theories, though, have something in common; they both describe qualitative changes in processing as the inevitable result of practice. Moving between either declarative to procedural, or algorithmic to retrieval processes, can be described as a qualitative change in processing.

It is against this background that Haider and Frensch (1996, 1999a, 1999b) have put forward their information reduction (IR) hypothesis. The central theme of their proposal was that over practice people learned to isolate task-relevant from task-redundant information. The irrelevant or redundant information was subsequently ignored leading to improved task performance. This occurred without any instruction over which segment of the task may be redundant, so the reduction strategy was learned incidentally. They (Haider & Frensch, 1996, 1999a, 1999b) claimed that this reduction in processing after practice was a general part of skill acquisition. If IR is a general part of skilled performance this raises

questions for two of the main theoretical approaches in attention and learning. Both Anderson's (1981) Act\* and Logan's (1988) instance theory, described qualitative changes in processing as skills are acquired. Neither account for the quantitative change proposed by the IR hypothesis. To reiterate, In instance theory the *way* information is processed changes. In contrast, as IR takes place, *what* information is processed changes. In instance theory the method changes from algorithm to retrieval, but the same information is processed, whereas in IR the amount of information processed is reduced. Neither Logan's nor Anderson's accounts of learning account for this quantitative change. This points to some shortcomings in both theoretical approaches and provides the motivation for an additional investigation of IR. Thus, the aim of the thesis was to further determine the role of IR within such contemporary theories.

Haider and Frensch's (1996, 1999a, 1999b) position was that not just qualitative but also quantitative changes occur with practice. Accepting this account, when performing a skill, individuals may not just employ retrieved examples or procedures, they may also reduce the information to which they attend. This must potentially be an important property of skilled task execution and so requires further study. The implication that quantitative changes occurred, necessitates an understanding of how these changes might occur. How do we come to selectively ignore elements deemed task-irrelevant? It is also important to understand how quantitative changes mesh with qualitative changes, and whether individuals vary in terms of the extent to which such quantitative changes occur.

Results suggest that IR was under conscious control and occurred at an early perceptual level (Haider & Frensch, 1999a, 1999b). IR also did not appear to be stimulus-specific

since it transferred across different stimulus-sets (Haider & Frensch, 1996). This latter point provided particular difficulties for an instance account of learning, since reduction was applied to novel items that could not be retrieved from memory. This research has provided a description of the properties of information reduction, particularly within the alphabet arithmetic task. IR though, by its very description, suggests that attention is given to one part of the stimulus, while withdrawn from the rest, thus research on attention becomes relevant. IR is also a ‘task strategy’; as a result the strategy selection literature is also important. So as Logan (2002) prefers, an understanding of this effect must span different main areas of research. The literature review spanning the areas of attention, learning and skill, helped to identify an additional sequence of research questions which were then addressed in the thesis.

## **8.1 RATIONALE FOR THE STUDIES**

To further investigate the IR hypothesis, the studies in the thesis manipulated different intrinsic task attributes of either an alphabet arithmetic task (AAT) or a target search task (TST). To summarise the studies;

### **8.1.1 EXPLORING INFORMATION REDUCTION USING THE AAT AND RANDOM LETTER-CASE.**

The first study of the thesis replicated the early results of Haider and Frensch (1996) using the AAT. Letter-case was randomised in the strings to introduce some noise into the stimuli. Even though letter-case was randomly varied, reduction still occurred by 500 trials. Interestingly, a number of participants reported the use of mnemonics: that is, they learned the triplets and so could avoid the calculation. It seems then, that within the AAT, retrieval

of triplet instances did take place and this was not disrupted by the random letter-case used for the strings.

#### 8.1.2 SEARCHING FOR INFORMATION REDUCTION USING A NEW TASK.

The second study investigated a new TST task in two separate experiments. The TST differed from the AAT in that each string was unique so no learning of instances through repetition was possible. Also the strings did not break down into alphabet calculation and recognition sub-tasks. IR was found to develop over practice in both these studies, but was more pronounced when the length of the strings was increased with additional letters.

#### 8.1.3 EXAMINING THE EXTENT TO WHICH TASK ATTRIBUTES MIGHT INFLUENCE INFORMATION REDUCTION.

The third study of the thesis manipulated three different task attributes within one TST experiment; these were source segmentation, task difficulty and string repetition. Compared to a standard condition which had none of these properties, there was some evidence that IR increased for each of the three 'attribute' conditions. The most unequivocal result was for the repetition condition, in which the number of unique string instances was reduced. For this condition, regression slopes came closer to zero and test errors for irregular strings containing novel target positions were also the highest.

During the experiment exploring task attributes, data were also collected to allow an investigation of individual differences and IR. Such differences may possibly explain the performance range that was found within IR tasks. These differences occurred in each experiment conducted so far: the experiments seemed to contain a sub-set of participants

that did not reduce. The question was, did this group differ along some measurable psychometric? It was thought anxiety per se might be associated with a propensity for IR. Also the literature suggested that cognitive ability might moderate strategy selection. The hypothesis was that those high in anxiety or matrix reasoning aptitude, would be more likely to reduce.

To test this, a version of the state-trait anxiety index (STAI, Speiburger, 1969) and a sub-test of the Wechsler abbreviated scale of intelligence (WASI, Wechsler, 1999) were administered. Reducers and non-reducers were categorised from the data and compared with the results of these tests. Neither reported anxiety nor matrix reasoning tests successfully discriminated between those who reduced and those that did not. The only reliable measure to correlate with the level of reduction was participants self-report taken during the end-of-session questionnaire. This suggested that individuals were consciously aware of the rule when they reduced.

#### 8.1.4 INVESTIGATING THE INFLUENCE OF RULE REGULARITY ON INFORMATION REDUCTION.

The final study of the thesis began with a pilot study to investigate the effect of providing feedback at the end of each AAT block, rather than after each trial. This change seemed to lower the extent of reduction that occurred with this task. To regain the full IR effect required participant instructions stressing speed as necessary whilst verifying each string. The main study, again using end-of-block feedback, varied the regularity of AAT strings at three different levels. Regularity here means how consistently information was relevant or irrelevant. For the 100% regular condition, during training the triplet was always relevant to



task completion, while the post-triplet was always irrelevant – never contained an alphabetic error. For the other two conditions regularity was set at 90% and 60%, so these training strings contained some post-triplet errors.

IR progressed as expected in the 100% condition, slopes dropped over blocks and irregular *test* errors were higher than regular during the test phase. IR did not occur for the 60% condition. This again was expected, since this condition was in effect a continuous test block for every block. The 90% condition appeared equivocal, slopes did not significantly reduce over blocks, but there was an increase in irregular test errors scores. More alphabetic errors were missed when they were irregular (post-triplet) than when these errors were within the triplet. This suggests at some level, post-triplet items were being ignored; there was a degree of IR for the 90% regular strings.

## **8.2 GENERALISING THE TASK**

As can be determined from this brief re-iteration of the results, the empirical goals of the thesis were concise: which task attributes moderate IR? The theoretical aims were more diverse. To begin with, one of the specific aims for this research was to address whether IR was a general part of skilled performance. Considerable literature has already been cited to support the view that people do seek strategies as they practice a task (for example Cheng, 1985; Doane, Alderton, Sohn, & Pellegrino, 1996). One thing that becomes clear from the literature, though, is that often these strategies are very domain specific. For example, they are to do with compass navigation (Newton & Roberts, 2000) or skill at a particular game (Shapiro & Raymond, 1989). The strategy formed for one domain will be no use at all in

the other. While it can be said that strategies in general, of which IR is one, are part of skill acquisition, it is a different thing to say IR is often a component of general skilled performance. To accept this, IR cannot be domain specific and must occur in many different tasks that include irrelevant information.

The stance taken here and by Haider and Frensch (1996) was that IR is likely to generalise to many tasks. To support their claim, they ran numerous experiments using the AAT with various manipulations (Haider & Frensch, 1996, 1999a, 1999b, 2002, in press). One reason for continuing with the AAT, was the difficulty in realising different paradigms that were both novel to the participant, and also subject to an IR strategy. However, attributes particular to the AAT may have been necessary for IR to develop. This was certainly a limitation with the earlier research. It is possible that the presence of irrelevant information may not alone be sufficient for an IR strategy to develop; certain task features may also be required. In order to claim IR is a general part of skill, it must therefore be demonstrated in other tasks that have different characteristics to the AAT.

The results from the TST described briefly above provide just this evidence. IR was found to occur in two separate slightly different experiments in Study 2. However, the effect was more convincing when the strings were increased to 13 items in length. The TST differs from the AAT conceptually in a number of ways. Most importantly, each TST string was unique, so did not repeat in a block. This tends to rule out a strict instance account as the basis of improvement. This was particularly important since the first study of the thesis found participants were forming mnemonics of the triplet instances in the AAT.

This agrees with earlier results that show IR to transfer across stimulus-sets (Haider & Frensch, 1996). However, this evidence alone did not rule out the necessity of instance learning during the development of IR. Indeed some repetition may have been necessary for any implicit mechanism to process the regularity of relevant information positioning. Later when the strategy was explicit and at a verbal level, transfer over different forms would be likely, as found in the transfer experiment (Haider & Frensch 1996, Expt 3). Finding IR with the TST meant that instance learning was not necessary (or possible) either during the development of the reduction strategy or for its deployment. Study 2 not only generalises IR to a new task, but also indicates recall of string instances was not necessary for the effect. This places IR further outside theories of skill acquisition reliant on episodic accounts, particularly Logan's (1988) instance theory.

One caveat with accepting this, is that current theories such as Logan's (1988), do not entirely prescribe what constitutes an 'instance'. If, as discussed during the literature review, something of the *method* by which a task solution is reached is also encoded or abstracted across instances, then this wider-ranging scheme cannot be ruled out at this point. However, Logan's contemporary instance theory (Lassaline & Logan, 1993; Logan, 1988, 1998, 2002) does not really encompass this type of learning, so still does not explain the type of performance improvements described during information reduction.

A second property of AAT that became apparent: was that the task split into two sub-tasks; alphabet calculation and alphabet recognition. The triplet required calculation, but the post-triplet just required recognition of that part of the alphabet, an over-learned task. There was no particular theory that indicated IR would be reliant on this, but there was some literature

concerning psychophysics that report similar cueing effects. That is, with some cue to indicate an area of relevance in a visual display, this information was used to reduce to it. It seems then, that when cued in such studies, people do reduce the information to which they attend (for example, Chun and Jiang, 1998, 1999; Olson and Chun, 2001). The separate sub-tasks of the AAT provide a strong cue of separate sources. Green and Wright's (2003) study indicates that reduction occurs too whichever of these sources was processed first. Thus, it would seem such separation may be important for IR. Again the TST completely removes such a possibility, since there was just one holistic task; to search the whole string for targets. So IR can progress even if the stimulus is not pre-defined into relevant and irrelevant sources.

Therefore, thesis Study 2 provided evidence that IR could occur across two different tasks. This was important as it suggested that IR may generalise beyond the AAT. This increases confidence that information reduction may be a general part of skilled performance. The study also ruled out the possibility that specific task attributes of the AAT were necessary for reduction to occur. This is not to say, though, that task attributes could not moderate the effect.

### 8.2.1 TASK ATTRIBUTES

Accepting that strategies are often formed to aid task completion; the next question is what particular conditions moderate the adoption of these methods: and IR in particular? The use of any strategy may be more or less likely depending on task conditions. Schunn (2001) reports that strategy use may be adaptive: in particular his participants seemed to moderate their strategy depending on the success of that method. Turning to IR, Haider and Frensch

(1999a) found IR was more or less employed depending on task instruction to be speeded or accurate, respectively. Slopes did reliably decline for accuracy conditions, but not to the same extent as for speeded conditions. Thus, it seems the reduction strategy may be moderated by different variables. It is not a case of reduction or not, the level may vary along a continuum. In this case, the level of reduction depended on an *explicit* task instruction. Logically, it was not surprising that IR increases with speed pressure: the goal of reduction must be either to reduce task load or increase the speed by which it is accomplished. The need to work quickly must heavily weight this goal.

The relation between speed pressure and the main reasons to develop a reduction strategy may mean that this was the primary variable that will moderate IR. However, it is also possible that other more *intrinsic* task variables will also moderate reduction. Study 3 of the thesis addressed this point. Of particular interest was the role of stimulus familiarity. For other strategies, Schunn et al. (1997) has indicated that the choice between calculation and retrieval methods depends on a feeling of familiarity with the test items. This was interesting, since in Study 1, a number of participants realised the AAT strings were repeating and did appear to learn the triplets. Comparing Study 1 and Study 2 in the thesis also highlights the possibility. First, IR was more complete in the AAT than the TST, and, second, that the triplet repetitions may have been the cause of the difference.

The most unequivocal result from Study 3 was that repetition of the relevant string segments within blocks had an effect on IR. The regression slopes came closer to zero and the number of test block errors for irregular strings were highest for this group. Also response errors reached a similar level compared to the AAT in Study 1. Overall, it seems

repetition did affect IR, and probably played a role in previous studies. Possibly, repetition within the strings meant that the post-triplet items were not given much attention at all during the test block. The parsimonious explanation is that participants realised the repetition within the strings, so were confident not to check the whole instance. As a result, not only were many of the test alphabetic errors missed, but the strategy persisted through the test block even following error feedback. Applying this to the account of Schunn et al. (1997), when the number of unique string instances were limited to a learnable set, a sense of string familiarity or a 'feeling-of-knowing' resulted in greater adoption of an IR strategy. Of course, string familiarity or a feeling-of-knowing could be re-conceptualised as confidence in strategy deployment. The more familiar the strings seem in some way, the greater the confidence to reduce what is processed.

What of the other task attributes tested; segmentation of relevant from irrelevant items via brackets and an increase in task difficulty via an alphabet calculation? There was some indication that increasing difficulty heightened reduction over the standard condition (that contained none of the additional attributes). However, the results were not entirely convincing. Perhaps the task did not differ enough from the standard condition for this, though it could be imagined that too much difficulty could prevent IR developing, rather than encouraging it. As Newton and Roberts (2000) report, often strategy formation depends on the availability of enough free resource to develop a new method. The effect of task difficulty therefore requires further investigation. Segmenting the two string sources, to cue the relevant area, also reduced regression slopes over standard, but did not increase the error rates of interest.

Overall, it appears that intrinsic task attributes as well as extrinsic instruction can moderate IR. Particularly it seems some perceived repetition in the task can change the level of reduction. However, none of these attributes were necessary for reduction to take place; since it also occurred in the standard condition. Rather, the differing task attributes served to moderate the effect.

### 8.2.2 IR HEURISTIC?

IR appears to be moderated by both extrinsic and intrinsic task attributes. This suggests that the strategy may be affected by variables other than just speed pressure. An important question though is: how consistent or regular a divide between relevant and irrelevant information is needed for reduction to develop? It is possible that any mechanism that detects this regularity needs the stimulus to be entirely consistent for reduction to take place, as was the case in the previous thesis Experiments 1-4. On the other hand, any level of such regularity may trigger reduction. There were two reasons to suppose IR would be sensitive to source regularity. Generally, Shunn (2001) reports that participants adapt strategy selection depending on success rate. More specifically, when triplet position was randomised across the AAT strings (Haider & Frensch, 1999b, Expt 1), reduction was attenuated.

However there is also a great deal of literature indicating that people utilise heuristic strategies. One key feature of these is that they are often 'rules of thumb', used even if they do not always lead to a correct solution (Gigerenzer & Selten, 2001). If IR was employed to cope with speed stress or reduce processing load, it can be imagined that under these

conditions it could be used even if it sometimes led to errors in performance. Thus, it was an empirical question whether IR would be adopted even if it was a sub-optimal strategy. The final study of the thesis investigated this very interesting aspect of IR. In the study the positional regularity of relevant information was held at 100% within the triplet, 90% or 60%. In the latter two conditions alphabetic errors occurred 10% and 40% of the time in the post-triplet segment respectively.

IR was found at 100% regularity and not at 60%, which suggested that IR did not occur when the strategy could not be applied to a large proportion (40%) of the strings. However the 90% condition was more interesting. For this condition while regression slopes did not decline reliably over blocks, during the test phase there were more post-triplet alphabetic errors missed than triplet errors. Since a fatigue explanation has already been ruled out, the most likely reason for this difference was that less attention was paid to the post-triplet region, so processing had reduced. This supports the view that even when it was not the optimal strategy, IR may still be employed 'at a level': that is, IR may be a heuristic strategy, though one that is adapted depending on success rate. In more naturalistic situations with more practice at a task, even if it is a sub-optimal strategy, IR may be even more likely. This is a somewhat different interpretation of the very similar results reported by Haider and Frensch (in press), a view which is expanded further in the next section.

### **8.3 FATE OF UNATTENDED ITEMS AND RESIDUAL PROCESSING**

Eclectically taking the evidence provided by Frensch, Haider, Runger, Neugebauer, and Werg (2003), Haider and Frensch (2002, 1999b) and also Haider, Frensch and Joram (in



press) their overall proposal appears to be that IR progresses as a switch-like process of strategy selection. In re-analysing their data, they point to individual discontinuities in response times over practice (Frensch, et al., 2003; Haider & Frensch, 2002, Haider, Frensch & Joram, in press). They take these step changes to indicate the strategy switch for the individual: that is, when IR was adopted. Thus, at some stage during practice, the suggestion was that a sudden decision was made not to process the irrelevant items. This happens so consistently that a sudden and continuing drop in response time occurred. It seems we are asked to believe that one moment there was no reduction and then suddenly it was adopted. The gradual decline in RT and regression slopes found in previous studies (Haider and Frensch 1996, 1999a, 1999b) arose when, as they say, all the discontinuities at individually different times were averaged across subjects (Haider & Frensch, 2002).

In one of the earlier studies (Haider and Frensch, 1999b), evidence was provided that IR occurred at an early perceptual level. Using eye tracking equipment they determined that following reduction, participants did not fixate irrelevant stimuli. So it appears a decision was taken early on to pay attention to only relevant items. It is doubtful that Haider and Frensch believe that irrelevant items were never fixated following reduction or the RT discontinuity, but overall a clear-cut decision to suddenly reduce appeared to be the take-home message. In other words, either reduction had taken place or it had not.

### **8.3.1 ARE STRATEGY SHIFTS CAUSED BY DATA-DRIVEN OR VOLUNTARY PROCESS: CRITIQUE OF HAIDER, FRENCH AND JORAM (IN PRESS).**

This approach was central to the argument in their latest study (Haider, Frensch & Joram, in press). In the final experiment of the paper they manipulated the regularity of the

redundancy in a very similar way to the last study of the thesis. The post-triplet items were either always, 90% or 60% redundant. So for example in the 90% condition, alphabetic errors occurred within the triplet 90% of the time and post-triplet for 10% of the strings. Their aim did *not* focus on whether IR developed even when information was not consistently redundant. Rather they were interested in the voluntary nature of the strategy: was it a conscious choice or the inevitable result of practice? To test this each block contained some speeded trials. Their rationale was that the strategy may be learned even if the redundancy was not entirely consistent. However, it may not be employed if it could not be applied to the whole range of problems. They further hypothesised that if the strategy was under conscious control, it may still be used under lower levels of regularity if there was some additional pressure to do so. In this case, speed pressure was used to encourage IR.

This was exactly what they seemed to find when the regularity was 90% consistent (Haider, Frensch & Joram, in press). Under no speed pressure in this condition, regression slopes did not decline. Whereas they did for the 100% condition in which post-triplet items were always redundant. Also, when 100% and 90% regularities were compared, more post-triplet alphabetic errors were missed at test for the 100% condition. So there was a marked difference between these two conditions under no speed pressure. In comparison, during speeded trials the difference in performance between these two groups disappeared. A rather complicated measure was used to support this. Ad hoc time 'windows' were decided that were gradually shorter and shorter, then the percentage error rate over practice for each increasingly strict window was calculated. They found that both over practice and for each

window, error rate was very similar for both 100% and 90% conditions, but much higher for the 60% regularity.

The only explanation for this was that a reduction strategy had been learned and used in both the highly regular conditions under speed pressure, but this was not available to the 60% regular group. Since the strategy was unavailable to this latter group, it led to more errors under the time constraint. The important point they suggest, was that the difference between the two highly regular conditions found during the normal trials, disappears under speed pressure. Thus the strategy was available for the 90% group but was only chosen when there was a need to. Otherwise reduction was avoided, since it could not be applied across all the strings. This, they suggest, revealed the voluntary use of a reduction strategy that was not the inevitable result of practice.

This was a compelling argument, but one that rests on accepting that IR did not occur for the 90% condition when there was no speed pressure. Unfortunately their results are not clear-cut, and focused mostly on the regression slopes. As they state, these did not decline for either the 90% or 60% levels, indicating that no reduction had taken place. However, if the test block errors are investigated, it seems a different result surfaces. Of particular interest were errors rates for irregular test strings, in which the alphabetic-error occurred post-triplet. While these errors were reliably highest for the 100% condition, the 90% condition was also statistically higher than the 60% condition. For the 90% condition during the test block, there were 31.0 mean errors for these irregular strings but only 3.71 for the regular strings. It is difficult to think of any other reason than IR for this difference.

If this was the case, an IR strategy was also adopted at a level, when there was no speed pressure and the strategy was not successful for all the problems

The results of both Haider, Frensch & Joram (in press) and the final thesis study, were the same, but the interpretation differed. Contrary to their proposal, it is argued here that information reduction did occur when error position regularity was held at 90%. Haider, Frensch & Joram (in press) provided some additional analysis that suggested participants in the 90% condition behaved more like the 60%. This they felt supported the view that those in the 90% were not adopting the strategy entirely. Allowing this, there was still the interesting result that the string-length effect did not reduce for the 90% group, while the test error scores do indicate reduction. To resolve this paradox returns us to an earlier point; is IR a switch like process or a strategy adopted at different levels? That is, IR could be suddenly and completely adopted, or the redundant area could be intermittently checked at varying frequencies.

## **8.4 EXPLAINING INFORMATION REDUCTION**

### **8.4.1 RESIDUAL PROCESSING HYPOTHESIS**

The stance taken here is that IR was adopted at different levels. Or to put it another way, some level of what will be termed ‘residual processing’ of irrelevant items was present even following the strategy switch. Further, this level varies depending on a number of task conditions. There are a number of threads of evidence to support this view. Throughout each experiment in the thesis an analysis of residual processing has been approached. This was achieved by comparing regression slopes over blocks for target-absent strings

compared to target-present strings. For example, once an alphabetic error was found the validation search would end. Since the errors always followed the brackets, there should be no hint of a string-length effect: that is, regression slopes for these strings should represent complete reduction in the task. This is generally what was found: the target present slopes were close to zero.

Within the analysis, then, as the target absent and present slopes came together, this indicated less processing of irrelevant items. The difference between the two string types therefore, indirectly measured the level of irrelevant item processing over blocks. Haider and Frensch (1996, 1999a, 1999b) do not carry out this particular comparison, as understandably their concern was to provide proof that IR developed, and not highlight the degree to which it did not. However, this aspect of IR also informs about the mechanism of such a strategy.

The thesis studies, especially study 3, indicated residual processing was moderated by various task attributes. This suggested that IR was not an all-or-nothing strategy, rather it was adopted at different levels. It could be said this was expected, since the string-length effect was seen to reduce gradually over blocks as the level of reduction increased. However, Haider and Frensch (2002) hypothesised that this gradual change resulted from averaging performance across participants. In comparison, individual data suggested the switch to IR was sudden, leading to discontinuities in performance. This then still describes a switch to the new strategy. Normally in the literature of strategy selection, once discovered a strategy was completely adopted until it was no longer successful. IR appears

to be at odds with this account. The reliable difference between standard and reduced instances conditions in study 3, strongly indicated that it was adopted to different degrees.

An alternate suggestion is that residual processing was really just an artifact of the varying string lengths. For example, the relevant segment becomes more difficult to locate as the strings increase in length. However, an unexpected result from the study in the previous chapter removes this argument. In the 100% condition within this study, the redundancy was always regular. So in effect this was a replication of the standard IR task. Not surprisingly IR was found according to the slope and error criteria. Also as before, for the test block slopes increased, again suggesting a return from IR. So, for the test block participants began processing the whole string again. Importantly though, feedback was at the end of a block, so the only way post-triplet errors could have been detected during the test phase was by some level of residual processing. The redundant segment must have been checked to notice the change in redundancy. No indication was given that the test block commenced, so the conclusion is that the whole string was always checked 'at some level' of frequency throughout.

This was perhaps the best evidence that residual processing occurred during the task, even following practice. There is no doubt though that the frequency of checking the irrelevant items became very low, otherwise irregular test errors would not be missed at all. One can imagine a situation, though, where the residual processing was enough to sustain the string-length effect, but the checking was not frequent enough to catch all the test errors. This explains both Haider, Frensch & Joram (in press) results and those found here in the last experiment. When the regularity of the redundancy was reduced to 90%, residual

processing was high enough to sustain some string-length effect so the slopes did not reduce significantly over blocks. However, this checking of the post-triplet string items was not frequent enough to catch the higher level of post-triplet alphabetic errors during the test phase. This describes how both no reduction in slopes over practice could be found along with significantly higher misses for post-triplet than triplet alphabetic errors.

Accepting the above points reveals IR as a strategy adopted at different levels. When IR was adopted at lower levels, residual processing of irrelevant string items increased. The level of reduction may be changed by any number of task parameters, such as stimulus familiarity, time pressure or success rate. It also seems that some level of residual processing can be viewed as endemic to the strategy, at least within the confines of lab based training levels. That residual processing is common would be expected, an attentional mechanism that ignores part of the visual scene entirely would not be likely to evolve. Just because a tiger was never in a particular part of the forest, does not mean the area should never be scanned for danger. In reality, it is likely to be scanned, but at a reduced level compared to common locations for threats. As Sekular and Blake (1994, p.264) write, while reading, something scurrying in the periphery instantly captures attention. Thus, attenuated or in the case of IR, residual processing, may be a prevalent part of the attentional mechanism, a property that allows ignored areas to re-capture attention when necessary.

#### 8.4.2 IR AS VIGILANCE

To further support the claim that residual process may be a common part of IR requires some re-conceptualisation of the tasks. To attend constantly to a relevant segment while

attempting to ignore irrelevant items, to some extent, suggests a type of vigilance task. There was a great deal of research in this area from the 1940s to the 1960s, particularly to do with radar operator performance (Mackworth, 1948; Mackworth & Kaplan, 1964). In such tasks, attention to a relevant area tends to decrease over time and targets are missed: that is, attention slips away from the task in hand (Davies & Parasuraman, 1981). Broadbent (1964, p.120) also suggests a sort of internal blink as one reason for the vigilance decrement over time.

This blink was an integral part of his well known filter theory (Broadbent, 1964). The filter it seems has a bias for channels that have not recently been attended. When attending to one area (or channel) over time, this bias will mean that other areas are momentarily selected. To use his phrase, “attention wanders”. Later proponents of his filter theory also point to ‘attentional slippage’, in order to account for the effect of irrelevant stimuli (Lachter, Forster & Ruthruff, 2004). It is possible, then, that attentional slips to unattended irrelevant stimuli may be a common aspect of attention in general and also the IR strategy. Whether this is a conscious decision or the property of an automatic attentional process requires further research.

## **8.5 INTEGRATING QUALITATIVE AND QUANTITATIVE ACCOUNTS OF CHANGES THAT OCCUR DURING SKILL ACQUISITION.**

### **8.5.1 TOWARDS INTEGRATION**

Information reduction has been found across various task manipulations and with different task types. As such, it is likely to be a wide reaching strategy found in many tasks that have some element of redundancy. Subsequently, any theory of skill acquisition must integrate



IR within its bounds. As written previously, eminent theories such as Logan's (1988) instance theory and Anderson's (1981) ACT theory respectively capture instance- and procedure-based accounts of learning. Both describe mainly qualitative changes and so do not fully account for the quantitative changes brought about by IR. The question is can these theories, in principle; accommodate a process whereby for a certain type of task, only relevant task information is selected.

One possible view is that IR interacts with one of the schemes mentioned. For example, a reduction strategy may form a certain type of procedure; Anderson, Matessa and Douglass (1995) suggest 'information gathering' could be seen as a production rule within the ACT theory. Similarly, Logan and Etherton (1994) have shown that attention guides what 'gets into' an instance. IR could be an attentional mechanism that guides the content in this way. Both these points were previously well made by Haider and Frensch (1999b). For the latter point though there does seem to be further evidence presented in the thesis. In Study 3 the most convincing change in IR was brought about by reducing the set-size of unique strings. However, it was the way the set-size was reduced that is interesting here.

To re-iterate, the task was to detect, within each string, a target from a memory set of three, then to decide for target present strings whether letters either side came before or after it in the alphabet. Crucially, the target was always second from left and there were only ten target and letter pair combinations presented for this condition. Important for this discussion, the letters following the letter-target-letter triplet were generated at random. Because of this only the number of unique 'triplets' were reduced in number. The number

of unique strings when viewed holistically remained high. So set-size could only become an advantage as reduction progressed.

Therefore, the superior performance for this group must have been achieved by learning and retrieving only the triplet. If, alternately, the whole string had been retrieved, the set-size would have been much greater and the manipulation would have made no difference. This supports the conclusion that IR directed the content of the string instances, which was constrained to the first three items. In other words, the learned instance was conditional on attention, a point repeated by Logan (2002, p. 396) with the premise that everything is conditional on attention. Even within his instance theory of attention and memory (ITAM) model, he emphasises two model parameters that turn up the attentional gain on desired objects and turn down the gain on undesired ones.

#### *Random-walk theory and IR*

Random walk theory has been introduced into this thesis at two points. Firstly Haider and Frensch have invoked these models as a partial explanation of how task instructions may effect IR at the level of random walk decision processes. Secondly, Nosofsky and Palmeri (1997) and also Palmeri (1997), have extended Logan's instance theory using a random walk process in order to explain similarity gradients in categorisation tasks. It is interesting that both accounts of IR and instance theory refer to random walk models (RWM) in some way. At first thought, this may suggest some scope to integrate the two at the level of a random walk decision processes. Including a random-walk process would certainly allow for some quantitative as well as qualitative change in processing within instance theory.

The main problem with this is that the reduction of information within a random walk (RW) is not the change in processing described in IR. The change in information processing in a RW concerns the *number* of comparisons before a response, whereas the change following IR concerns *which* comparisons are made. This argument was more fully developed earlier in the thesis (p221-230).

Undoubtedly, incorporating RW into instance theory as described by Nosofsky and Palmeri (1997) does allow for one type of quantities change in information processing. Such a model allows for fewer evidences or exemplars to be retrieved before a response. However, where this explanation, and instance approaches in general, fall down, is accounting for performance improvements that do not seem to rely on instance retrieval. The standard target search task (TST, Study 2) is a case in point. Each string was unique, so the performance improvement was mainly due to IR alone. The change in behaviour did not need to invoke any current instance account. If, for example, Logan's theory is to incorporate learning that includes the selective use of information, then the concept of 'an instance' will have to be relaxed. An instance will have to include some analogous knowledge that allows the similarities across instances to be represented.

IR may be an attentional strategy that guides the content of an instance, but as it stands Logan's instance theory (1988) does not really model this type of attentional constraint. Even the later ITAM model (Logan, 2002) does not fully specify an attentional mechanism that leads to a reduction in task processing. The exemplar-based random walk model proposed by Nosofsky and Palmeri (1997) does allow for *similar* instances of an exemplar to be retrieved, but any abstraction across these is not yet specified in the model. There

appears to be some work required before IR can be fully integrated into contemporary theories of skill acquisition.

#### **8.5.2 A STAGE FRAMEWORK**

An alternative approach is to propose that each theory or hypothesis describes discrete stages during learning. Anderson, Fincham & Douglass (1997), suggest four discrete, if overlapping, stages of cognitive skill. In the first stage, novel problems were initially solved by analogy with previously learned exemplars. Eventually some declarative rules may be discovered and applied. With further practice production rules developed (the procedural embodiment of the declarative rules). Finally, as the novel problems were repeated, straightforward retrieval of examples from memory could answer the problems. This described solving problems by analogy, procedures and instance retrieval, all of which may happen in one task. One could imagine adding strategy development generally, and IR specifically, to this list. There certainly appears to be evidence that IR and instance retrieval occur within the alphabet arithmetic task (AAT). Participants report the use of mnemonics and reducing the set-size to a learnable number also increases reduction. It is possible that a full account of skilled performance will require a synthesis of theories of the sort described by Anderson and colleagues.

#### **8.6 THE MECHANISM OF IR: IMPLICIT OR JUST INCIDENTAL?**

At some point during both the AAT and the TST, participants realised that part of the string was irrelevant to the task. How this redundancy was noticed though is open to question. It

is difficult to avoid positing some sort of mechanism that abstracts the regularity (target position) across episodes (string instances). So at some level an instance explanation is required, but how was the regularity represented? It is at this point that an account of IR seems to cross into the domain of implicit learning. In essence implicit learning describes a process whereby some regularity in the environment was learned incidentally and without awareness, yet affected behaviour. There is though some debate over the existence of implicit phenomena, especially when learning is conceptualized as the abstraction of rules (Cock, Berry & Gaffan, 1994; Stadler et al., 2000). Putting this to one side, though, it seems that the underlying regularity of error position was learned *incidentally* and this did lead to IR.

It is difficult to avoid the assumption that instances were encoded in some way. How else could the similarity between them become noticed? The actual instances are less important here, though, than the commonality between them. There is though some question over how this commonality may be represented. Rule abstraction at an implicit level is not currently a popular notion. The current theory is that some sort of abstract analogy forms (Brooks & Vokey, 1991; Huddy & Burton, 2002; Vokey & Brooks 1992, 1994) or similarity processing comes into play (Cock, Berry & Gaffan, 1994). Under these viewpoints reduction would occur for a new string either because it was analogous to some abstracted form or because it was similar to fragments of other instances contained in memory. Some incidental-learning mechanism could use either of these forms to represent the regularity.

The word 'incidental' in the previous paragraph was chosen carefully, since incidental and implicit learning are not the same. There is little doubt that the regularity in both the AAT

and TST was learned incidentally, since no instruction was given to this end and the tasks did not require the strategy to be learned. However there is some question over the implicit nature of the knowledge. The working definition of implicit learning is that knowledge is learned without intention. Further, whilst the individual is not consciously aware of this knowledge it nevertheless effects behaviour (Kelly, Burton, Kato & Akamatsu, 2001; Mathews, 1995; Perruchet, Chambaron & Ferrel-Chapus, 2003; Reber, 1993). In each experiment in the thesis though, a significant percentage of the participants were able to verbalise the regularity found within the strings. This would mean that the knowledge was explicit, not implicit.

Perhaps, though, this is only true towards the end of practice; initially implicit knowledge could guide behaviour. Again this does not seem to be the case: Haider and Frensch (2002, In press) provide evidence otherwise. Not only do they report a sudden discontinuity in performance when the strategy was adopted, but also at this point the knowledge of the strategy can be verbalised. Given that implicit learning should affect performance before explicit awareness of the regularity develops, a change in behaviour before this point would be expected. IR seemed to take place at the same time as the regularity was realised consciously. Haider and Frensch (2002) do not seem to have found any evidence of implicit learning affecting behaviour; participants were aware of the rule at the time performance changed. Only detailed re-analysis of their results, or a future experiment, will determine if there is any gradual reduction before the discontinuity in response times. At present the results do seem to suggest an incidental, but not implicit, learning mechanism, one that results in an explicit verbal rule.

### *Knowledge representation*

The way that task-relevant and task-redundant knowledge is represented generally is also an interesting question. For example, does information reduction change the qualitative representation of task relevant information following practice? Perhaps the converse is true, qualitative changes in knowledge representation lead to IR, whichever, the ways in which qualitative and quantitative changes, such as IR, mesh is a fascinating area. Both this and the exact nature of the mechanism that detects environmental regularities pose important future empirical questions.

## **8.7 PRACTICAL IMPLICATIONS: REAL-WORLD TASKS**

The practical implications of research for real-world situations must be a concern for experimental cognitive psychology. There is no doubt that the goal in understanding skill acquisition is to find worthwhile applications for everyday life. Understanding IR should align with this goal, since information reduction may play a part in general skilled performance. For example, IR during high pressure complex perceptual tasks may differentiate the novice from the expert, and may also explain the cause of some performance errors. However, IR needs to be understood at a cognitive rather than just a descriptive level.

The investigation so far, has revealed IR to be a strategy that may be a general part of skilled performance. Already it has been shown specifically in two separate task types, the AAT and TST. Additionally, it appears to occur to some extent in more complex tasks, such as air traffic control simulations (Lee and Anderson, 2000). Experienced pilots also

seem to reduce the information selected from their display panel (De Maio, Parkinson, Leshowitz, Crosby & Thorpe, 1976). These examples strongly suggest that IR is not just a very specific laboratory based phenomena, but one that is important for real-world skills.

#### 8.7.1 ERROR AVOIDANCE

What are the implications of this? One point to make here is that IR may have both positive and negative consequences. When task loads are high or there is significant time pressure, IR may be one way of managing the situation. Learning to ignore irrelevant information will free resources to process crucial information and so should improve both the speed and accuracy of performance. On the other hand though, redundant information may not remain redundant throughout all situations. If this happens then an IR strategy may lead to error prone performance.

This is born out by the experimental data: first, under increasing time pressure a reduction strategy becomes more likely (Haider & Frensch, 1999a) so it is one way of dealing with time constraints. Second, all the studies find that when the regularity changes and targets appeared in untrained positions, error rates increased for these. So there is also a downside to reduction. As Haider and Frensch (1996, p. 335) suggest, faulty information reduction may be one cause contributing to accidents and for this reason alone experimentation on IR is required. The Three Mile Island incident described in the literature review is a case in point: inappropriate reduction of attended information in this event almost led to a disaster of huge proportion.



### 8.7.2 CONDITIONS FOR IR

For similar reasons to those given above, the conditions under which IR is facilitated and inhibited warranted further study. The thesis lays out findings that showed that not only extrinsic factors such as task instruction moderate reduction; intrinsic task attributes also had an affect. Haider and Frensch (1999a) have shown that instructions to be accurate or work quickly change the degree of IR. The results of the thesis, especially Studies 3 & 4 provide evidence that attributes of the task itself can also moderate the strategy. In particular, it seems IR increases when some repetition of the stimulus was apparent. This means that tasks that contain both some redundancy and also a level of repetition are very likely to result in information reduction.

Other attributes, such as cues discriminating relevant from irrelevant information and overall task difficulty also showed some signs of moderating the effect. Study 4 reported here, indicated that IR may occur even if it sometimes was not the optimal strategy. Reduction developed even though it could not be applied across all the problems. So IR may be adopted, even as a sub-optimal strategy and this seems even more likely if there is some speed pressure (Haider, Frensch & Joram, in press). This represents a first understanding of the conditions under which IR might occur. All these factors should be taken into consideration when designing tasks, depending whether reduction is appropriate or not. One very obvious application of the results would be to ensure that important, but not always relevant information is grouped in displays near to information sources that need to be constantly checked. In this way inappropriate reduction may be avoided. It seems logical that this is considered in the design of displays. There may yet be more

practical uses for an understanding of IR, in the training of novices, in accident avoidance and in the layout of instrument panels.

## **8.8 IMPLICATIONS FOR FUTURE RESEARCH**

The IR effect replicates in a number of experiments and tasks run by different researchers, so appears to be a real effect. However the hypothesis is falsifiable; when the regularity was less than 60% reliable IR did not seem to occur for the alphabet arithmetic task (AAT). It was however a robust effect that did not rely on large group numbers to reach reliable differences. So, it seems IR lends itself well to experimentation. It is an important topic not just because it is outside contemporary skill theories, but because it may have very valid implications for understanding real-world performance, operator errors, and devising more effective instructional methods. Perhaps, though, what makes it most interesting and takes it beyond just another strategy, is that it is not task-specific. It is a phenomenon that may be a general part of skill learning and expert performance.

IR was a robust effect, though there did seem to be individual variation in task strategies. In each experiment there was a sub-set of people who never reduced. In one way this shows that reduction was not a trivial occurrence, it was not a method that was very obvious in the tasks. Additionally, it also suggests that there are some interesting individual differences to investigate. The first attempt at this in chapter 6, did not meet with a great deal of success. The measures chosen did not distinguish between the sub-groups of reducers and non-reducers. However there are many other personality and intelligence measures that may

prove more fruitful in this respect. This would be one interesting area of future research. Such an understanding of personality variables that affect reduction may also be useful in selecting appropriate personnel for particular jobs. For instance, if thorough checking for safety reasons is needed, then perhaps individuals prone to reduction should be avoided.

The attention literature seems to set location as a particularly important driver of attentional selection (Driver, 2001; Lamy 2001). Since IR is an attentional strategy, it is possible that the location of relevant information is the prime factor in the strategy. It is possible though that 'context cues' may also be used to decide when to reduce or not. Instructions to be speeded or accurate certainly appear to change the context of a task and moderate IR. It is possible that some other background variable could also be used to decide whether to process the whole stimulus or not. Experimentally this could be a background context such as 'to check the whole string when the screen background is light, but only check the triplet when the background is dark'. The work of Green and Wright (2003) suggest conjunctions of relevant information from different sources may be learned in this way, but further research is necessary. It is also possible that factors other than location may guide reduction, such as coherent motion of a sub-set of the scene or the colour or shape of relevant and irrelevant stimulus segments.

A number of intrinsic task attributes have been shown to moderate reduction, such as repetition and the regularity of target position. There are potentially many other attributes to investigate. Within the thesis state/trait anxiety was measured, but did not correlate with reduction. However, it was found the task did not really increase anxiety levels, and there was no manipulation of anxiety within a group. With no situational stress high trait anxiety

would not lead to state anxiety. So it is too early to rule out the effect of increased anxiety on reduction. As with speed pressure, anxiety may work in the same way to increase reduction. Overall, there appear to be many areas of information reduction yet to explore.

## 8.9 CONCLUSIONS

The thesis provides evidence that IR occurs in different tasks, an important point considering that much of the previous research repeatedly used versions of the alphabet arithmetic task. From a theoretical perspective this strongly suggests that IR is a general part of skilled performance for tasks that have redundant elements. A further implication is that this reduction in processing following practice needs to be incorporated in any overarching theory of skill acquisition.

Integration with theories such as Anderson's ACT\* (1981) and Logan's (1988) instance theory proves difficult at present. The main difficulty is that neither allows for a mechanism that reduces the amount of information processed. One possible mechanism suggested was a random-walk process. This has been used to explain both similarity gradients and IR itself. However, while *quantitative* processing changes can be described within random-walk models, these are not the strategic changes within IR. An alternative is to view IR as a mechanism that guides attention and so helps determine what is learned during practice. It seems though that integration is still a long way off. More work is required, Logan (2002, p.392) views modeling IR as a type of learning within his framework (and indeed others), as an important direction for future research.

The literature often suggests a strategy switch, indeed the later work of Haider and Frensch (2002) and Haider, Frensch and Joram (in press) describe the sudden adoption of the IR strategy, a discontinuity in performance. The stance taken for the thesis, though, was that information reduction takes place at different levels. This was the basis of the residual-processing hypothesis put forward: there tends to be some level of residual processing of the irrelevant items. In other words it is not a question of whether the strategy is adopted or not, but rather the level or frequency to which it is used.

A further contribution towards understanding the conditions which facilitate and inhibit reduction has been achieved. The thesis provides findings that not only extrinsic, but also intrinsic task attributes moderate reduction. This moderation takes the form of different levels of residual processing as hypothesised. Understanding IR in this way provides some synthesis for the conflicting reports between the final study of the thesis and Haider, Frensch and Joram's (in press) most recent work.

The final conceptualisation of IR, then, is to describe it as a general-purpose intentional and attentional strategy, one that improves performance following practice, by constraining processing to only task-relevant elements. However, reduction is not usually complete; rather, some checking of the irrelevant items occurs, though this residual processing may be modified by any number of task variables. Practice may not really make perfect, but it normally does lead to performance improvements. One way that the gain was achieved here was through the reduction in task processing, which Haider and Frensch (1996) term as information reduction.

## REFERENCES

- Anderson, J.R. (1981). *Cognitive Skills and Their Acquisition*. Hillsdale
- Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369–406.
- Anderson, J.R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J.R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, Vol.94, No2, 192-210.
- Anderson, J.R. (1992). Automaticity and the ACT\* Theory. *American Journal of Psychology*, Vol. 105, No. 2, 165-180.
- Anderson, J.R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*. Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, J.R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, (4), 932-935.
- Ashby, F.G. (1987) Counting and timing models in psychophysics and the conjoint Weber's law. *Journal of mathematical psychology*, 31,4, 419-428
- Avant, L.L. & Thieman, A.A. (1985), On Visual Access to Letter Case and Lexical/Semantic Information. *Memory and Cognition*, Vol. 13, (5), 392-404.
- Ball, K., and Sekuler, R. (1981) cues reduce direction uncertainty and enhance motion detection. *Perception & Psychophysics*, 30, 119-128
- Barnard, P.J. (1985). Interacting cognitive subsystems: A psycholinguistic approach to short term memory. In A. Ellis (Ed.), *Progress in the Psychology of Language*, Vol. 2, London: Erlbaum.
- Berry, D.C., & Dienes, Z. (1993). *Implicit Learning: Theoretical and empirical issues*. Hove, England: Erlbaum.
- Besner, D. (1983), Basic Decoding Components in Reading: Two dissociable Feature Extraction Processes. *Canadian Journal of Psychology*, Vol. 37, (3), 429-438.
- Blalock, H.M. jr. (1971) *Causal models in the social sciences*. Macmillan: London.

- Bliss, J.P. & Acton, S.A. (2003) Alarm mistrust in automobiles: How collision alarm reliability affects driving. *Applied Ergonomics*, 34, 499-509.
- Bornstein, R.F. (1989), Exposure and Effect: Overview and Meta-Analysis of Research. 1968-1987, *Psychological Bulletin*, Vol .106, No. 2, 265-289.
- Bransford, J.D., Sherwood, R.D., Vye, N.J. & Rieser, J. (1986) teaching thinking and problem solving. *American Psychologist*, 41, 1078-1089
- Bright, J.E.H., & Burton, A.M. (1994). Past midnight: Semantic processing in an implicit learning task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 47A, 71-89.
- Bright, J.E.H., & Burton, A.M. (1998). Ringing the changes: Where abstraction occurs in implicit learning. *European Journal of Cognitive Psychology*, 10, (2), 113-130.
- Brigman, S. & Cherry, K.E. (2002). Age and skilled performance: contributions of working memory and processing speed. *Brain and Cognition*, 50, 242-256.
- Brill, J.C., Mouloua, M., Hancock, P.A., and Gilson, R.D. (2004) A multi-sensory protocol for evaluating workload associated with new army Technologies. 24<sup>th</sup> *Army Science Conference*, Nov. 29<sup>th</sup>, Orlando Florida.
- Broadbent, D.E. (1958). *Perception and Communication*. Pergamon Press LTD.
- Broadbent, D.E. (1964). *Perception and communication*. Macmillan: New York
- Broadbent, (1990). Effective decisions and their verbal justification. *Philosophical Transactions of the Royal Society of London*, Series B, 327, 493-502.
- Bryan, W.L., Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review* 6, 345-375.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, Vol. 97, No. 4, 523-547.
- Campitelli, G., & Gobet, F. (2005) The mind's eye in blindfold chess. *European Journal of Cognitive Psychology*. Vol 17(1), 23-45.
- Cara, F., and Lagrange, V. (1999). Emerging expertise in process control. *Ergonomics*, 42 (11), 1418-1430.
- Carmody, D.P., Nodine, C.F., and Kundel, H.L. (1980) An analysis of perceptual and cognitive factors in radiographic interpretations. *Perception*, 9, 339-344.
- Cheng, P.W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, Vol. 92, No. 3, 414-423.

- Chun, M.M. and Jiang, Y (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28-71.
- Chun, M.M. and Jiang, Y (1999) Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360-365.
- Cock, J. J., Berry, D.C., & Gaffan, E.A. (1994). New strings for old: The role of similarity processing in an incidental learning task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 47A, 1015-1034.
- Cohen, A.L. & Nosofsky, R.M. (2003) An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, 47, 150–165.
- Cole, R.A. & Haber, R.N. (1980), letter to the Editor: Reaction Time to Letter Name or Letter Case. *Acta Psychologica*, Vol. 44, 281-285.
- Croner, L.J. and Albright, T.D., (1994). Segmentation by color improves performance on a visual motion task. *Investigative Ophthalmology and Visual Science Supplement* 35, 1643.
- Croner, L.J. and Albright, T.D., (1997). Image segmentation enhances discrimination of motion in visual noise. *Vision Research* 37, 1415–1427.
- Crossman, E.R.F.W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, 2, 153-166.
- Davies, D.R. & Parasuraman, R. (1981). *The psychology of vigilance*. Academic Press Inc. London.
- Delaney, P.F., Reder, L.M., Staszewski, J.J., & Ritter, F.E. (1998). The strategy specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, 9, 1-7.
- De Maio, J., Parkinson, S., Leshowitz, B., Crosby, J., & Thorpe, J.A. (1976). *Visual Scanning: Comparisons between student and instructor pilot*. (AFHRL-TR-76-10, AD-A023 634). Williams Air force Base, AZ: Air Force Human Resources Laboratory, Flying Training Division.
- Department of Employment. (1971). *Glossary of Training terms*. 2<sup>nd</sup> edition. London:HMSO.
- Deutsch, J.A. & Deutsch, D. (1963). Attention: some theoretical considerations.



*Psychological Review*, 70, 80-90.

- Diaz, D.D., Hancock, P.A., and Sims, V.K. (2004) Ecological visual search under noise stress: Present and future research. *24<sup>th</sup> Army Science Conference*, Nov. 29<sup>th</sup>, Orlando Florida.
- Diaz, D.D., Sims, V.K., and Hancock, P.A. (2002) Memory for ecologically valid objects during visual search for two identical targets. *14<sup>th</sup> Annual Meeting of the American Psychological Society*, New Orleans Louisiana.
- Doane, S. M., Alderton, D. L., Sohn, Y. W., & Pellegrino, J. W. (1996). Acquisition and transfer of skilled performance: Are visual discrimination skills stimulus specific? *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1218-1248.
- Doane, S., Sohn, Y. W., & Schreiber, B. (1999). The role of processing strategies in the acquisition and transfer of a cognitive skill. *Journal of Experimental Psychology: Human Perception and Performance*, 25(5), 1390-1410.
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92, 53-78.
- Duncan, J. (1980). The locus of interference in the in the perception of simultaneous stimuli. *Psychological Review*, 87, 272-300.
- Easterbrook. J.A. (1959) The effect of emotion on cue utilization and the organisation of behavior. *Psychological Review*, 66, 183-201.
- Edwards, M. and Badcock, D.R., (1994). Global motion perception: interaction of the ON and OFF pathways. *Vision Research* 34, 2849–2858
- Edwards, M. and Badcock, D.R., (1996). Global-motion perception: interaction of chromatic and luminance signals. *Vision Research* 36, 2423–2432
- Erikson, B.A. & Eriksen, C. W. (1974). Effects of Noise Letters Upon the Identification of a Target Letter in a Non-search Task. *Perception and Psychophysics*, 16, 143-149.
- Eriksen, C.W. & St.James, J.D. (1986). Visual attention within and around the field of focal attention: A Zoom Lens model. *Perception and Psychophysics*, 40 (4), 225-240
- Eriksen, C.W. & Yei-Yu Yeh, (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 11,

No. 5, 583-597.

- Eysenck, M. W. (1979). Anxiety, learning and memory: A reconceptualisation. *Journal of Research and Personality*, 13, 363-385.
- Eysenck, M. W. (1982). *Attention and Arousal: Cognition and Performance*. Berlin: Springer.
- Eysenck, M. W. (1985). Anxiety and cognitive-task performance. *Personality and Individual Differences*, 6, 579-586.
- Eysenck, M. W. (1988). Anxiety and attention. *Anxiety Research*, 1, 9-15.
- Eysenck, M.W. (2000). *Psychology: A Students Handbook*. Psychology Press.
- Faul, F. & Erdfelder, E. (1992). GPOWER: A-priory, post-hoc and compromise power analysis for MS-DOS (Computer Program), Bonn, FRG; Bonn University, Dep of Psychology.
- Fitts, P.M. (1962). Factors in complex skill learning. In Glaser, R. (Ed) *Training Research and Education*, University of Pittsburg
- Fitts, P.M. & Posner, M.I. (1967). *Human Performance*. Belmont, CA: Brooks Cole.
- Frensch, P.A., Haider, H., Runger, D., Neugebauer, U., Voigt, S.V., and Werg, J. (2003) Verbal report of incidentally experienced environmental regularity: The route from implicit learning to verbal expression of what has been learned. in Jimenez, L. (Ed), *Attention and implicit learning*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Gibson, J.J. & Gibson, E.J. (1955) Perceptual learning: Differentiation or enrichment? *Psychological Review*, 62, 32-41
- Gibson, B.S., Li, L., Skow, E., Brown, K., and Cooke, L. (2000) Searching for one versus two identical targets: When visual search has a memory. *Psychological Science*, 11, 324-327.
- Gigerenzer, G & Selten, R. (2001) *Bounded rationality: The adaptive toolbox*. Cambridge, The MIT press.
- Gigerenzer, G & Todd, P.M. & ABC research group (1999) *Simple heuristics that make us smart*. Oxford University Press.
- Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1-40.
- Green, A.J.K. & Wright, M.J. (2003). Reduction of task relevant information in skill

- acquisition. *European Journal of Cognitive Psychology*, Vol. 15, No. 2, 267-290.
- Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, 30, 304–337.
- Haider, H., & Frensch, P. A. (1999a). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, 5(2), 129-151.
- Haider, H., & Frensch, P. A. (1999b). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 172–190.
- Haider, H. & Frensch, P. A. (2002) Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmieri (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*. Vol 28(2), 392-406.
- Haider, H., Frensch, P. A., & Joram, D. (pre-print). Are strategy shifts caused by data-driven processes or by voluntary processes? *Consciousness and Cognition*. xxx, xxxx-xxxx.
- Hancock, P.A., Szalma, J.L., and Weaver, J.L. (2004) The distortion of perceptual space-time under stress. . 24<sup>th</sup> Army Science Conference, Nov. 29<sup>th</sup>, Orlando Florida.
- Hintzman, D.L. (1986) “Schema abstraction” in a multiple-trace memory model. *Psychological review*, Vol. 93, No. 4, 411-428. *Journal of experimental psychology: Learning, memory and cognition*. Vol 17, No. 2, 196-209.
- Howell, D.C. (1997). *Statistical Methods for Psychology*. Fourth edition, Duxbury press.
- Huddy, V., & Burton, A.M. (2002). Generate and test: An alternate route to knowledge elicitation in an implicit learning task. *The Quarterly Journal of Experimental Psychology*, 55A, (4), 1093-1107.
- Hutchinson, J.M.C. & Gigerenzer, G. (2005) Simple heuristics and rules of thumb: where psychologist and behavioural biologist might meet. *Behavioural Processes*, 69, 97-124.
- Jimenez, L. (2002) *Attention and Implicit Learning*. John Benjamins Publishing Company: Amsterdam/Philadelphia

- Johnson, S. (1989). Inside TMI: Minute by Minute,  
<http://www.wowpage.com/tmi/main.html>
- Kahneman, D., Slovic, P. & Tversky, A. (1982) *Judgment under uncertainty: Heuristics and biases*, Cambridge university press.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology-Monograph*, 74, 657-690.
- Kelly, S.W. Burton, A.M., Kato, T., & Akamatsu, S. (2001). Incidental learning of real world regularities. *Psychological Science*, 12, (1), 86-89.
- Kim, J. and Ferree, G.D (1981) Standardization in causal analysis. In Marsden, P.V. (ed.) *Linear models in social research*. Sage publications: Beverly Hills: London.
- Klapp, S., Boches, C.A., Trabert, M.L. & Logan, G.D. (1991) Automatizing alphabet arithmetic: II. Are there practice effects after automaticity is achieved. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 17, No. 2, 196-209.
- Lachter, J., Forster, K.I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, Vol 111. No 4, 880-913.
- Lamberts, K., Brockdorff, N. & Heit, E. (2003) Feature-Sampling and Random-Walk Models of Individual-Stimulus Recognition, *Journal of Experimental Psychology: General*, Vol. 132, No. 3, 351–378.
- Laming, D.R.J. (1968) *Information theory of choice-reaction times*, Academic press.
- Lamy, D. & Tsal, Y. (2001) On the status of location in visual attention. *European Journal of Cognitive Psychology*, 13 (3), 305 – 342.
- Lassaline, M.E. & Logan, G.D. (1993) Memory-based automaticity in the discrimination of visual numerosity. *Journal of experimental psychology: Learning, memory and cognition*. Vol. 19, No. 3, 561-581.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 451-468.
- Lavie, N. (2000). Selective attention and cognitive control: Dissociating attentional functions through different types of load. In S. Monsell & J. Driver (Eds), *Control of Cognitive Processes: Attention and Performance XVIII*,

- Cambridge, MA: MIT Press.
- Lavie, N. & de Fockert, J.W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception and Psychophysics*, 65, 202-212.
- Lavie, N., Hirst, A., de Fockert, J.W. and Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, Vol. 133, No. 3, 339-354.
- Lavidor, M., Ellis, A.W. & Pansky, A. (2002) Case alternation and length effects in lateralised word recognition: studies of English and Hebrew. *Brain and Cognition*, 50, 257-271.
- Lee, F.J. & Anderson, J.R. (2000). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology*, 42, 267–316
- Lee, J.D. & See, K.A. (2004) Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50-80.
- Lehman, R.S. (1995) *Statistics in the behavioural sciences: A conceptual introduction*. Brooks/Cole: CA.
- Lewis, R.L. (2001) *Cognitive theory, Soar*. In International Encyclopedia of the Social and Behavioral Sciences. Amsterdam: Pergamon (Elsevier Science).
- Lieberman, H.R., Bathalon, G.P., Falco, C.M., Georgelis, J.H., Morgan III, C.A., Niro, P., and Tharion, W.J.(2004) The “Fog of War”: Documenting cognitive decrements associated with the stress of combat. . 24<sup>th</sup> Army Science Conference, Nov. 29<sup>th</sup>, Orlando Florida.
- Lincourt, A.E., Hoyer, W.J., and Cerella, J. (1997) *Selective Attention in the Acquisition of instance-based Learning: Adult Age Differences*. Manuscript submitted for publication, Referenced in Haider, H., & Frensch, P. A. (1999b). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 172-190.
- Lincourt, A.E., Rybach, J.M. & Hoyer, W.J. (1998). Aging, working memory and the development of instance-based retrieval. *Brain and Cognition*, 37, 100-102
- Lindsay, D.S. (1990) Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 1077-1083.

- Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Logan, G.D. (1998) What is learned during automatization? II. Obligatory encoding of spatial location. *Journal of experimental psychology: Human perception and performance*. Vol. 24, No. 6, 1720-1736.
- Logan, G.D. (2002). An instance theory of attention and memory. *Psychological Review*, Vol.109, No. 2, 376-400.
- Logan, G.D. & Etherton, J.L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 20, No. 5, 1022-1050
- Logan, G.D., Taylor, S.E. & Etherton, J.L. (1999). Attention and automaticity: Toward a theoretical integration. *Psychological Research*, 62, 165-181.
- Lovett, M.C. & Anderson, J.R. (1996) History of Success and Current Context in Problem Solving: Combined Influences on Operator Selection. *Cognitive Psychology* 31, 168–217.
- Mackworth, N.H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*. 1, 6-21.
- Mackworth, NH & Kaplan, IT. (1964). Eye movements during vigilance. *Perceptual & Motor Skills*. 18(2), 397-402.
- Marsden, P.V. (1981) *Linear models in social research*. Sage publications: Beverly Hills: London.
- Mathews, R.C. (1995). Implicit learning: Attacking the authoritarian view of mind. *Contemporary Psychology*, 40 (6), 525-528.
- Mayall, K., Humphreys, G.W. & Olson, A. (1997), Disruption to Word or Letter Processing? The Origins of Case-Mixing Effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 23, No. 5, pp 1275-1286.
- McGeorge, P., and Burton, A.M. (1990). Semantic processing in an incidental learning task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 42A, 597-609.
- McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, 332, 154-155.
- Miner, A & Reder, L.M. (1994). A new look at feeling of knowing: Its metacognitive role

- in question answering. In J. Metcalf & A.P. Shimamura (Eds.) *Metacognition: Knowing about knowing*. Cambridge, MA: Bradford.
- Mouloua, M., Gilson, R.D., Hancock, P.A., Brill, J.C., and Deaton, J. (2004) Development of effective decision strategies for alarm diagnosis in military aviation. 24<sup>th</sup> Army Science Conference, Nov. 29<sup>th</sup>, Orlando Florida.
- Murray, N. P., & Janelle, C. M. (2003). Anxiety and performance: A visual search examination of the processing efficiency theory. *Journal of Sport & Exercise Psychology*, 25, 171-187.
- Neisser, U. & Becklen, R. (1975) Selective looking: Attending to visually-specified events. *Cognitive Psychology*, 7, 480-494
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, Massachusetts: Harvard University Press.
- Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition*. Hillsdale, NJ: Erlbaum.
- Newton, E.J. & Roberts, M.J. (2000). An experimental study of strategy development. *Memory and Cognition*, 28, (4), 565-573.
- Niessen, C, Eyferth, K., and Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. *Ergonomics*, 42 (11), 1507-1520.
- Norman, D.A. & Shallice, T. (1980). *Attention to Action: Willed and Automatic Control of Behaviour*. (CHIP Report 99), San Diego, California,: University of California.
- Nosofsky, R.M. (1997) An Exemplar-Based Random-Walk Model Of Speeded Categorization And Absolute Judgment, in *Choice, Decision, and Measurement* A. A. J. Marley (Ed.). Lawrence Erlbaum Associates. NJ.
- Nosofsky, R.M. & Palmeri, T.J. (1997) An exemplar based random walk model of speeded classification. *Psychological Review*, 104, 266-300.
- Olson, I.R and Chun, M.M. (2001) Temporal contextual cueing of visual attention. *Journal of Experimental Psychology: Learning Memory & Cognition*, 27, 1299-1313.
- Pachella, R.G. & Pew, R.W. (1968). Speed-accuracy trade off in reaction time: effect of discrete criterion times. *Journal of Experimental Psychology*, 92, 378-384
- Palmeri, T.J. (1997) Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 324-354.

- Palmeri, T.J. (1997) An Exemplar-Based Random Walk Model of Perceptual Categorization. *Proceedings of the Interdisciplinary Workshop On Similarity And Categorisation*, University of Edinburgh, Scotland
- Palmeri, T.J. (1999), Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 543-550.
- Pask, G., & Scott, B. C. E. (1972). Learning strategies and individual competence. *International Journal of Man-Machine Studies*, 4, 217-253.
- Patrick, J. (1992). *Training: Research & Practice*. Academic Press.
- Payne, J.W., Bettman, J.R. and Johnson, E.J. (1993) *The adaptive decision maker*. New York : Cambridge University Press.
- Perruchet, P., Chambaron, S., & Ferrel-Chapus, C. (2003). Learning from implicit learning literature: Comment on Shea, Wulf, Whitacre, and Park. *The Quarterly Journal of Experimental Psychology*, 56A, (5), 769-778.
- Posner, M.I. (1967). Chronometric analysis of classification. *Psychological Review*, 74, 392-409.
- Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.
- Posner, M.I. & Keele, S.W. (1967). Decay of visual information from a single letter. *Science*, 156, 137-319.
- Posner, M.I. & Taylor, R.L. (1969). Subtractive method applied to separate visual and name components of multiletter arrays. *Acta Psychologica*, 30, 104-114.
- Rabbitt, P.M.A. (1989). Sequential reaction. In D.H. Holding (Ed.), *Human Skills*. (pp 153- 175), New York: Wiley.
- Rabitt, P. & Banerji, N. (1989) How Does Very Prolonged Practice Improve Decision Speed? *Journal of Experimental Psychology: General*, Vol. 118, No, 4, 338-345.
- Ratcliff, R. (1978) A theory of memory retrieval, *Psychological review*, 85, 2, 59-108
- Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, 88, 552-572.
- Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, Vol 92, No. 2, 212-225.
- Ratcliff, R. (1988) Continuous Versus Discrete Information Processing: Modeling Accumulation of Partial Information, *Psychological Review*, Vol. 95, No. 2, 238-255



- Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin and Review*, 9, 278-291.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical-decision task. *Psychological Review*, 111, 159-182.
- Ratcliff, R., & Hacker, M. J. (1981). Speed and accuracy of same and different responses in perceptual matching. *Perception and Psychophysics*, 30, 303-307.
- Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347-356.
- Ratcliff, R., & Rouder, J.N. (2000). A Diffusion Model Account of Masking in Two-choice Letter Identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 127-140.
- Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, Vol. 124, No. 3, 372-422.
- Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Reder, L.M. (1987). Strategy selection in question answering. *Cognitive psychology*, 19, 90-137.
- Reder, L.M. and Ritter, F.E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning Memory and Cognition*, 18, 435-451.
- Regeher, G., & Brooks, L.R. (1993) Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General*, 122, 92-114
- Rickard, T.C. (1997). Bending the power law: A CMPL of strategy shifts and the automatisisation of cognitive skills. *Journal of Experimental Psychology: General*, 126, 288-311.
- Roberts, M.J., Gilmore, D.J., & Wood, D.J. (1997). Individual differences and strategy selection in reasoning. *British Journal of Psychology*, 88, 473-492.
- Roberts, M. J., & Newton, E. J. (2001). Understanding strategy selection. *International Journal of Human-Computer Studies*, 54, 137-154.
- Ross, B.H. (1984) Reminders and their effects in learning a cognitive skill. *Cognitive*

- psychology*, 16, 371-416.
- Rouder, J.N. (2000) Assessing the roles of change discrimination and luminance integration: evidence for a hybrid race model of perceptual decision making in luminance discrimination. *Journal of Experimental Psychology: Human Perception and Performance*. Vol. 26, No. 1.359-378.
- Russo, M., Escolas, S., Sing, H., Thorne, D., Johnson, D., Redmond, D., Hall, S., Reed, W., Satiago, S., and Holland, D. (2004) Visual neglect in pilots in a simulated overnight flight. *24<sup>th</sup> Army Science Conference*, Nov. 29<sup>th</sup>, Orlando Florida.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1-66.
- Schneider, W., & Shiffrin, R. M. (1985). Theoretical note. *Psychological Review*, Vol. 92, No. 3, 424-428.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002) *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002) *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools Inc.
- Schunn, C.D. & Reder, L.M. (2001) Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology*, vol 130, No 1, 59-76.
- Schunn, C.D., Reder, L.M., Nhoyvanisviong, A., Richards, D.R. & Stroffolino, P. J. (1997), To calculate or not to calculate: A source activation confusion model of problem familiarity's role is strategy selection. *Journal of Experimental Psychology: Learning Memory and Cognition*, Vol 23, No.1, 3-29.
- Shanks, D.R., & St. John, M.F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367-447.
- Shapiro, K. L., & Raymond, J. E. (1989). Training of efficient oculomotor strategies enhances skill acquisition. *Acta Psychologica*, 71, 217-242.
- Shettleworth, S.J.(2005) Taking the best for learning. *Behavioural Processes*, 69, 147-149.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127-190.

- Siegler, R.S. & Jenkins, E.A. (1989) *How children discover new strategies*. Hillsdale, NJ: Erlbaum.
- Singleton, W.T. (1978). *The Analysis of Practical Skills*. MTP Press LTD.
- Smith, P.L. & Ratcliff, R. (2004) Psychology and neurobiology of simple decisions, *TRENDS in Neurosciences* Vol.27 No.3.
- Snowden, R. J., & Edmunds, R. (1999). Colour and polarity contributions to global motion perception. *Vision Research*, 39(10), 1813-1822.
- Sperling, G. & Doshier, B. (1986), Strategy and optimization in human information processing. In K.R. Boff; L. Kaufman & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol 1, New York: Wiley.
- Stadler, M.A., Warren, J.L., & Lesch, S.L. (2000). Is there cross-format transfer in implicit invariance learning? *The Quarterly Journal of Experimental Psychology*, 53A (1), 235-245.
- Sternberg, S. (1966, August 5<sup>th</sup>). High speed scanning in human memory. *Science*, 153, 652-654.
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 57, 421-457.
- Stewart, T. C. & West, R. L. (2006). Deconstructing ACT-R. In *Proceedings of the Seventh International Conference on Cognitive Modeling* (pp. 298-303). Trieste, Italy.
- Stolarz-Fantino, S. & Fantino, E. (2005) The rules we choose by. *Behavioural Processes*, 69, 151-153.
- Strayer, D.L. & Kramer, A.F. (1994a), Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning Memory and Cognition*, 20, 318-341
- Strayer, D.L. & Kramer, A.F. (1994b), Strategies and automaticity: II. Dynamic aspects of strategy adjustment. *Journal of Experimental Psychology: Learning Memory and Cognition*, 20, 342-365.
- Tiedemann, J. (1989). Measures of cognitive styles: A critical review. *Educational Psychologist*, 24(3), 261-275.
- Todd, P.M (2001) Fast and frugal heuristics for environmentally bounded minds. In Gigerenzer, G & Selten, R. (Eds) *Bounded rationality: The adaptive toolbox*. Cambridge, The MIT press.

- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242-248.
- Treisman, A.M. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.
- Treisman, A.M. & Geffen, G. (1967). Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology*, 19, 1-18.
- Treisman, A. M., & Gelade, G. (1980). A Feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136.
- Tukey, J.W. (1954) Causation regression and path analysis. In O. Kemp-Thorn et.al. (eds.) *Statistics and mathematics in biology*. Ames: Iowa State College Press.
- Underwood, G., Crundall, D., and Chapman, P. (2002) Selective searching while driving: The role of experience in hazard detection and general surveillance. *Ergonomics*, 2002, Vol. 45, No. 1, 1-12.
- Vokey, J.R., & Brooks, L.R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, No. 2 328-344.
- Walker, P. & Hinkley, L. (2003), Visual memory for shape-colour conjunctions Utilises structural descriptions of letter shape. *Visual cognition*, Vol. 10, (8), pp 987-1000.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence - WASI*. San Antonio: The Psychological Corporation, Harcourt Brace & Company.
- Welford, A.T. (1976). *Skilled Performance: Perceptual and Motor Skills*, Scott, Foresman and Company.
- Weltman, G., Smith, J. E., & Egstrom, G. H. (1971). Perceptual narrowing during simulated pressure-chamber exposure. *Human Factors*, 13, 99-107.
- Wickens, C.D., & Dixon, S.R. (2005) Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature. *Technical report AHFD-05-01/MAAD-05-01*, Aviation human factors division institute of aviation. University of Illinois
- Wickens, C.D., Gempler, K. & Morphey, M.E. (2000) Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation human factors*, 2, 99-126.

- Xu, W., Wickens, C.D. & Rantanen, E. (2004) Imperfect conflict alerting systems for the cockpit display of traffic information, *Technical report AHFD-04-8/NASA-04-2*, Aviation human factors division institute of aviation. University of Illinois
- Yerkes, R.M., and Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482
- Zizak, D.M. & Reber, A.S. (2004), Implicit preferences: The role(s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition*, 13, 336-362.

## **APPENDICES**

## Appendix 1: INFORMED CONSENT FORM

Informed Consent Form for a study on the effect of practice on performance. (IR Study 1)

### Information for Participants

The reason for running this experiment is to investigate the effect of practice at a task on its performance, both in speed and accuracy.

The task you will have to complete consists of verifying strings of letters to see if they follow the alphabet. These strings will be presented to you on a computer screen in sets of 100, you will be able to rest at the end of each set.

The experiment will take about one or two hours to complete, for which you will receive either £5 or £10. The time taken to do the experiment and the payment you receive are dependent on which group you have chosen.

Please note you are free to leave the experiment at any time if you feel you do not wish to continue. **If you have any other questions** please ask the experimenter now. Thank you for your participation, it is appreciated.

---

1. I, the undersigned, voluntarily agree to take part in the study investigating of the effect of practice on performance. I am over 18 years of age. To the best of my knowledge I am in good general health.

2. I have read the information for participants and understand what I will be expected to do. I have had the opportunity to ask questions, and my questions have been satisfactorily answered. I agree to the restrictions placed on me. I will be able to follow the procedures required of me, and agree to follow instructions as requested

3. **I am free to withdraw from the study at any time without the need to justify my decision.** Likewise, the investigators may dismiss me from the study without the need of justification.

4. I will not be referred to by name in any report concerning the study. I shall not claim to be entitled to restrict in any way the use to which the results of this study may be put. I will be fully protected in accordance with the Data Protection Act.

5. I am willing to take part in this study

Signed: \_\_\_\_\_ Date: \_\_\_\_\_

I have questioned the aforementioned subject to confirm their understanding of their participation in this study and now witness their signature above.

Witness \_\_\_\_\_ Date: \_\_\_\_\_

**Appendix 2: END QUESTIONNAIRES**

**FINAL QUESTIONNAIRE IR 1**

Have you done this type of task before? .....If yes what do you remember about it  
.....

Did you notice anything about the strings?  
.....  
.....  
.....

Did you notice anything about the errors?  
.....  
.....  
.....

Did you start to do the task differently as you went on – how?  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

Did you notice the errors always followed the digit in brackets?  
.....  
.....  
.....

Is there anything else you would like to say?  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....



**FINAL QUESTIONNAIRE IR 2, 3 & 4**

Have you done this type of task before? .....If yes what do you remember about it  
.....

Did you notice anything about the strings?  
.....  
.....  
.....

Did you notice anything about where the memory items appeared?  
.....  
.....  
.....

Did you start to do the task differently as you went on – how?  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

Did you notice the Memory-Set items were always in the second position from left?  
.....  
.....  
.....

Is there anything else you would like to say?  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

**FINAL QUESTIONNAIRE IRS**

Have you done this type of task before? .....If yes what do you remember about it  
.....

Did you notice anything about the strings?  
.....  
.....  
.....

Did you notice anything about where the errors appeared?  
.....  
.....  
.....

Did you start to do the task differently as you went on -- how?  
.....  
.....  
.....

Did you notice the errors tended always to follow the digit in brackets?  
.....  
.....  
.....

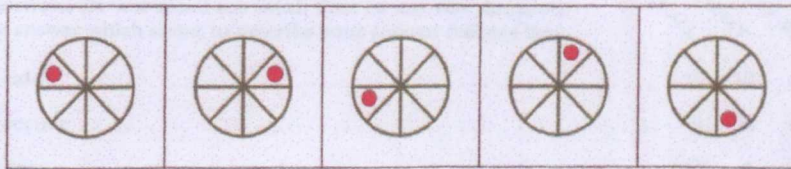
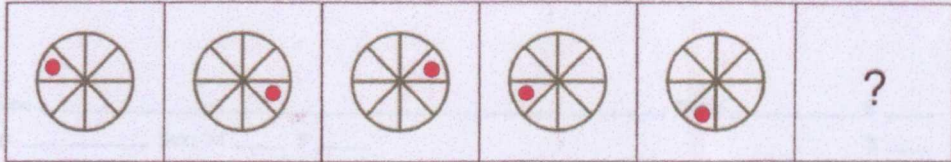
Is there anything else you would like to say?  
.....  
.....  
\*\*\*\*\*

Please indicate below on the scale how important in this task you think accuracy is  
versus speed by marking the line with a stroke?



### Appendix 3: EXAMPLE WECHSLER ABBREVIATED SCALE OF INTELLIGENCE (WASI) PROBLEMS

Select the missing figure from the five alternatives



1

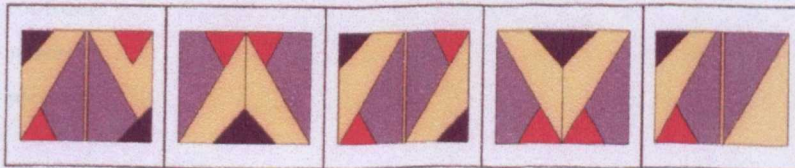
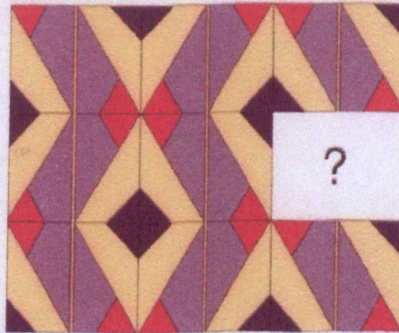
2

3

4

5

27



1

2

3

4

5

28

## Appendix 4: STATE-TRAIT ANXIETY INVENTORY (STAI)

### SELF-EVALUATION QUESTIONNAIRE

Name \_\_\_\_\_ Date \_\_\_\_\_ S \_\_\_\_\_  
Age \_\_\_\_\_ Sex: M \_\_\_\_\_ F \_\_\_\_\_ T \_\_\_\_\_

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then blacken in the appropriate circle to the right of the statement to indicate how you feel *right now*, that is, *at this moment*. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

	NOT AT ALL	SOMEWHAT	MODERATELY SO	VERY MUCH SO
1. I feel calm .....	①	②	③	④
2. I feel secure .....	①	②	③	④
3. I am tense .....	①	②	③	④
4. I feel strained .....	①	②	③	④
5. I feel at ease .....	①	②	③	④
6. I feel upset .....	①	②	③	④
7. I am presently worrying over possible misfortunes .....	①	②	③	④
8. I feel satisfied .....	①	②	③	④
9. I feel frightened .....	①	②	③	④
10. I feel comfortable .....	①	②	③	④
11. I feel self-confident .....	①	②	③	④
12. I feel nervous .....	①	②	③	④
13. I am jittery .....	①	②	③	④
14. I feel indecisive .....	①	②	③	④
15. I am relaxed .....	①	②	③	④
16. I feel content .....	①	②	③	④
17. I am worried .....	①	②	③	④
18. I feel confused .....	①	②	③	④
19. I feel steady .....	①	②	③	④
20. I feel pleasant .....	①	②	③	④



## SELF-EVALUATION QUESTIONNAIRE

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then blacken in the appropriate circle to the right of the statement to indicate how you *generally* feel. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe how you generally feel.

	ALMOST NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
21. I feel pleasant .....	①	②	③	④
22. I feel nervous and restless .....	①	②	③	④
23. I feel satisfied with myself .....	①	②	③	④
24. I wish I could be as happy as others seem to be .....	①	②	③	④
25. I feel like a failure .....	①	②	③	④
26. I feel rested .....	①	②	③	④
27. I am "calm, cool, and collected" .....	①	②	③	④
28. I feel that difficulties are piling up so that I cannot overcome them .....	①	②	③	④
29. I worry too much over something that really doesn't matter .....	①	②	③	④
30. I am happy .....	①	②	③	④
31. I have disturbing thoughts .....	①	②	③	④
32. I lack self-confidence .....	①	②	③	④
33. I feel secure .....	①	②	③	④
34. I make decisions easily .....	①	②	③	④
35. I feel inadequate .....	①	②	③	④
36. I am content .....	①	②	③	④
37. Some unimportant thought runs through my mind and bothers me .....	①	②	③	④
38. I take disappointments so keenly that I can't put them out of my mind .....	①	②	③	④
39. I am a steady person .....	①	②	③	④
40. I get in a state of tension or turmoil as I think over my recent concerns and interests .....	①	②	③	④

## Appendix 5 : GLOSSARY OF TERMS

Terms And Abbreviations	Key
IR	See Information Reduction
Information Reduction	Term first used by Haider and Frensch 1996 to describe their hypothesis that quantitative change in information processing may occur over practice
String-length	Number of items in a string
String-length effect	Increasing reaction time with increased number of string items
Triplet	Letter-digit-letter segment of an AAT string
Post-triplet	Those items following the triplet
Regularity	How consistent error position is in a string
Novel or untrained position	Used to refer to test targets appearing in positions different from the proceeding training blocks
Training blocks	Trials during which the task is practiced
Test block or test phase	Usually the final transfer block where target position changes
Target	Used interchangeably to refer to alphabetic errors or memory set targets in a string
RT	Reaction or response time
AAT	Alphabet arithmetic task
TST	Target search task
Regression slope	Regression coefficient in a regression analysis
Coefficient	See regression slope
Mean error rate	The mean number of errors in a block
Percentage error rate	Percentage of errors for a string type in a block
Practice	Repeated exposure with a task
Practice trials	First few trials to familiarise participant with the task
Residual processing	The degree to which redundant task items are processed
Strategy	A voluntary goal directed behaviour, usually to reduce effort or increase accuracy
Heuristic	Rule of thumb or sub-optimal strategy
Sub-optimal strategy	A strategy that does not use all the available information and so does not achieve the best possible result

p	Statistical Probability
Cognitive load	Cognitive task difficulty
Perceptual load	Perceptual task difficulty
Letter case	Whether letters are in upper or lower case – capitalised or not
Segment	Part of a stimulus string
Segmentation or perceptual segmentation	Visual cue separating the string into segments
Cue or cueing	Some task property that directs attention to some part of the stimulus
Reduction	See IR
Skill acquisition	Gaining proficiency at a task
Stimulus	In the experiments this refers to an alpha or alpha numeric sting presented sequentially via computer monitor
E-prime	Experiment generation software
ERTS	Experiment generation software
Visual Basic	High level computer programming language
PC	Micro Computer
Two-alternate forced choice task	Task where there are two possible responses and a response is compulsory to move on to the next task
Feedback	Indication of errors on a trial, may also include response speed. This may also be at the end of a block where mean error rate and response time is provided. Note error feedback may also be accompanied by a tone
Irrelevant	Refers to item or stimulus segment that does not need to be processed for a correct response
Redundant	While item or segment information is relevant to string verification, it is redundant if duplicated elsewhere.
Verification	Determine if a string is correct or not or contains a target or not
Demand characteristics or experimenter effects	Where questions or inferences expressed by the experimenter can influence participants answers or judgments.

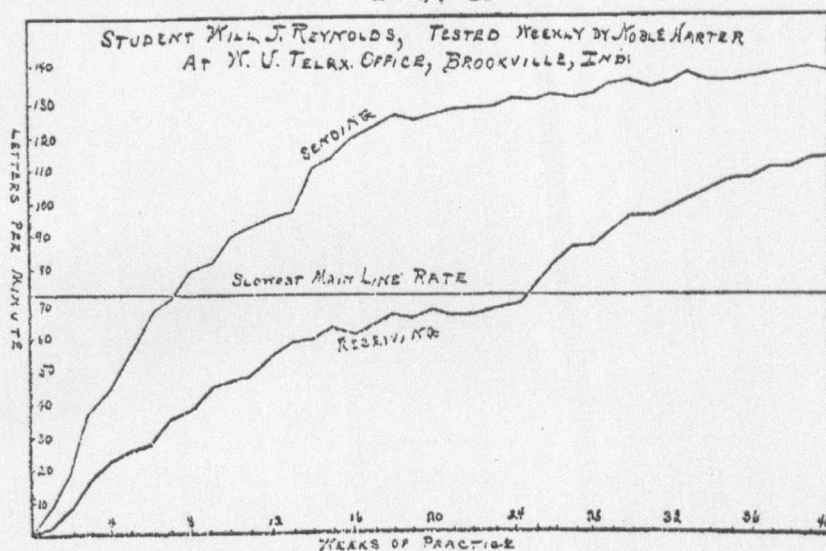


# Appendix 6 : Bryan & Harter (1899) Learning Curves.

350

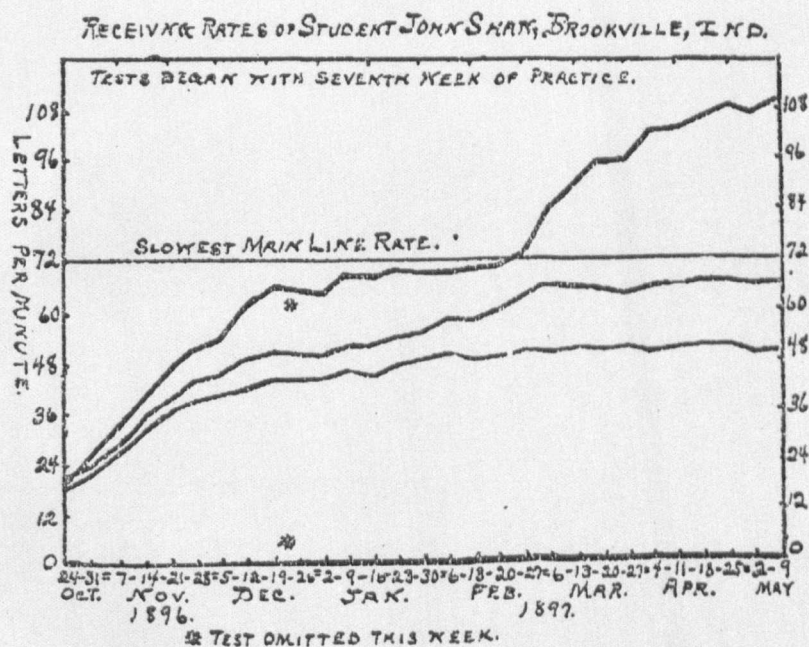
WILLIAM L. BRYAN AND NOBLE HARTER.

Fig. X.



Reproduced from PSYCH. REV., IV., 44.

Fig. XI.



Connected discourse curve at the top; word curve in the middle; letter curve at the bottom.

Supplied by The British Library - "The world's knowledge"